

NGT-01-002-099

NASA CONTRACTOR
REPORT

NASA CR-183553

RESEARCH REPORTS - 1988 NASA/ASEE SUMMER FACULTY
FELLOWSHIP PROGRAM

The University of Alabama
Tuscaloosa, Alabama
and
The University of Alabama in Huntsville
Huntsville, Alabama

December 1988

Final Report

(NASA-CR-183553) RESEARCH REPORTS: 1988
NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
Final Report (Alabama Univ.) 804 pCSCL 051

N89-21726
--THRU--
N89-21760
Unclass
0187249

G3/99

Prepared for
NASA, George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama 35812

RESEARCH REPORTS

1988 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

George C. Marshall Space Flight Center

The University of Alabama

and

The University of Alabama in Huntsville

EDITORS:

Dr. L. Michael Freeman
Assistant Professor of Aerospace Engineering
The University of Alabama

Dr. Charles R. Chappell
Associate Director for Science
Marshall Space Flight Center

Ms. Ernestine K. Cothran
University Affairs Officer
Marshall Space Flight Center

Dr. Gerald R. Karr
Chairman of Mechanical Engineering
The University of Alabama in Huntsville

1. REPORT NO. NASA CR-183553	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.	
4. TITLE AND SUBTITLE Research Reports - 1988 NASA/ASEE Summer Faculty Fellowship Program		5. REPORT DATE December 1988	
		6. PERFORMING ORGANIZATION CODE	
7. AUTHOR(S) Editors: M. Freeman, R. Chappell, E. Cothran, G. Karr		8. PERFORMING ORGANIZATION REPORT #	
9. PERFORMING ORGANIZATION NAME AND ADDRESS The University of Alabama Tuscaloosa Alabama and The University of Alabama in Huntsville		10. WORK UNIT NO.	
		11. CONTRACT OR GRANT NO. NGT-01-002-099	
12. SPONSORING AGENCY NAME AND ADDRESS National Aeronautics and Space Administration Washington, DC 20546		13. TYPE OF REPORT & PERIOD COVERED Contractor Report	
		14. SPONSORING AGENCY CODE	
15. SUPPLEMENTARY NOTES			
16. ABSTRACT <p>For the 24th consecutive year, a NASA/ASEE Summer Faculty Fellowship Program was conducted at the Marshall Space Flight Center (MSFC). The program was conducted by The University of Alabama and MSFC during the period June 6, 1988, through August 12, 1988. Operated under the auspices of the American Society for Engineering Education, the MSFC program, as well as those at other NASA centers, was sponsored by the Office of University Affairs, NASA Headquarters, Washington, D.C. The basic objectives of the programs, which are in the 25th year of operation nationally, are (1) to further the professional knowledge of qualified engineering and science faculty members; (2) to stimulate an exchange of ideas between participants and NASA; (3) to enrich and refresh the research and teaching activities of the participants' institutions; and (4) to contribute to the research objectives of the NASA centers.</p> <p>The Faculty Fellows spent 10 weeks at MSFC engaged in a research project compatible with their interests and background and worked on collaboration with a NASA/MSFC colleague. This document is a compilation of Fellows' reports on their research during the summer of 1988. The University of Alabama Report No. BER-452-94 presents the Co-Directors' report on the administrative operations of the program. Further information can be obtained by contacting any of the editors.</p>			
17. KEY WORDS Cryogenics, storm modeling, expert systems, tethered satellites, propulsion, atomic oxygen, superconductors, crystal growth, super computer optimization, control systems, gamma rays, hypervelocity impact, material properties, (software reliability), weld properties,		18. DISTRIBUTION STATEMENT CFD, (microgravity biology). Unclassified - unlimited. <i>Ernestine Cothran</i> 12/15/88 Ernestine Cothran University Affairs Officer	
19. SECURITY CLASSIF. (of this report) Unclassified	20. SECURITY CLASSIF. (of this page) Unclassified	21. NO. OF PAGES 830	22. PRICE NTIS

PREFACE

This document is a collection of technical reports on research conducted by the participants in the 1988 NASA/ASEE Summer Faculty Fellowship Program at Marshall Space Flight Center (MSFC). This was the twenty-fourth consecutive year the program has been conducted at MSFC. The 1988 program was administered by The University of Alabama (UA) in cooperation with MSFC and The University of Alabama in Huntsville (UAH). The program was operated under the auspices of the American Society for Engineering Education (ASEE) with sponsorship and funding from the Office of External Relations, NASA Headquarters, Washington, D.C. The MSFC program was one of ten such Aeronautics and Space Research Programs funded by NASA Headquarters in 1988. Similar programs were conducted at seven other NASA centers. The basic common objectives of the NASA/ASEE Summer Faculty Fellowship Program are:

- a. To further the professional knowledge of qualified engineering and science faculty members;
- b. To stimulate an exchange of ideas between participants and NASA;
- c. To enrich and refresh the research and teaching activities of participants' institutions; and,
- d. To contribute to the research objectives of the NASA centers.

The MSFC Faculty Fellows spent 10 weeks (May 31 through August 5, 1988) working with NASA scientists and engineers on research of mutual interest to the university faculty member and the NASA colleague. The editors of this document were responsible for selecting appropriately qualified faculty to address some of the many problems of current interest to NASA/MSFC. A separate document (UA-BER Report No. 452-94, September 1988) reports on the administrative aspects of the 1988 program. This document contains the technical reports on research performed by the individual 1988 participants. The NASA/ASEE program is basically a two-year program to allow indepth research by the university faculty member. The reports are arranged in alphabetical order.

TABLE OF CONTENTS

I.	Antar, Basil: University of TN Space Institute	"Cryogenic Fluid Management in Space"
II.	Baginski, Michael: Auburn University	"Finite Element Simulation of Thunderstorm Electrodynamics in the Proximity of the Storm"
III.	Elake, Jean: AL A&M University	"Refining, Revising, Augmenting, Compiling and Developing Computer Assisted Instruction K-12 Aerospace Materials for Implementation in NASA Spacelink Electronic Information System"
IV.	Bykat, Alex: University of TN at Chattanooga	"SIRE - a Simple Interactive Rule Editor for NICBES"
V.	Carrington, Connie: University of SC	"Small Expendable Deployer System Measurement Analysis"
VI.	Chang, Chia-Bo: TX Tech University	"The Role of Radiation-Dynamics Interaction in Regional Numerical Weather Prediction"
VII.	Chauhan, Rajinder: NC A&T State University	"Computer Codes for Thermal Analysis of a Solid Rocket Motor Nozzle"
VIII.	Chiu, Huei-huang: University of IL at Chicago	"Radiation Effect On Rocket Engine Performance"
IX.	Chyu, Mingking: Carnegie Mellon University	"A Performance/Loss Evaluation of SSME HPFTP Turbine"

- | | | |
|--------|--|--|
| X. | Eldridge, Hudson:
University of
Central AR | "X-Ray Fluorescence Sur-
face Contaminant Analyzer
A Feasibility Study" |
| XI. | Floyd, Stephen:
The University of
AL in Huntsville | "Intelligent Data Reduc-
tion for Autonomous Power
Systems" |
| XII. | Fromhold, Albert:
Auburn University | "Experimental Results On
Atomic Oxygen Corrosion
of Silver" |
| XIII. | Garland, Michael:
Memphis State
University | "Construction of an Al-
ternating Gradient Mag-
netometer" |
| XIV. | Hall, Kenneth:
MS State University | "Guidance for an Aero-
assisted Orbital Transfer
Vehicle" |
| XV. | Han, Samuel:
TN Technological
University | "Transient Natural Con-
vection Heat and Mass
Transfer in Crystal
Growth" |
| XVI. | Hornfeck, William:
MS State University | "Optimization of Large
Matrix Calculations for
Execution On The Cray X-
MP Vector Supercomputer" |
| XVII. | Irwin, Dennis:
OH University | "Investigation of Practi-
cal Applications of H
Control Theory to the
Design of Control Systems
for Large Space
Structures" |
| XVIII. | Kvale, Thomas:
The University of
Toledo | "The Measurement of Angu-
lar Differential Cross
Sections at the SSL
Atomic Scattering
Facility" |
| XIX. | Lestrade, Patrick:
MS State University | "Angular Response Cali-
bration of the Burst and
Transient Source Experi-
ment" |

- | | | |
|---------|--|---|
| XX. | McDonald, Gary:
The University of
TN at Chattanooga | "The Simulation of the
Alternate Turbopump De-
velopment High Pressure
Oxygen and Fuel Turbo-
pumps for the Space
Shuttle Main Engine Using
the Shaberth Computer
Program" |
| XXI. | McGuire, Stephen:
AL A&M University | "Fluctuation Analysis of
Relativistic Nucleus-
Nucleus Collisions in
Emulsion Chambers" |
| XXII. | Patterson, James:
FL Institute of
Technology | "Electron Mobility in
Mercury Cadmium Tellu-
ride" |
| XXIII. | Powell, Eugene:
GA Institute of
Technology | "Simulation of Water Va-
por Condensation on Lox
Droplet Surface Using
Liquid Nitrogen" |
| XXIV. | Schonberg, William:
The University of
AL in Huntsville | "Further Investigations
of Oblique Hypervelocity
Impact Phenomena" |
| XXV. | Springer, William:
University of AR | "Evaluate the Application
of Modal Test and
Analysis Processes to
Structural Fault
Detection in MSFC - STS
Project Elements" |
| XXVI. | Temple, Enoch:
AL A&M University | "SSME Propulsion Perfor-
mance Reconstruction
Techniques" |
| XXVII. | Tucker, Dennis:
GA Tech | "Dynamic Fatigue Testing
of Zerodur Glass-Ceramic" |
| XXVIII. | Uber, James:
The University of
AL in Huntsville | "Software System Safety" |
| XXIX. | Walsh, Daniel:
CA Polytechnic
State University,
San Luis Obispo | "The Keyhole Region in
VPPA Welds" |

- XXX. Whitaker, Kevin:
The University
of AL "The Role of the Turbu-
lent Prandtl Number in
Turbine Blade Heat Trans-
fer Prediction"
- XXXI. Yeager, Dorian:
The University
of AL "Development of a Proto-
type Commonality Analysis
Tool for use in Space
Programs"
- XXXII. Young, Ronald:
The University of
AL in Huntsville "Techniques for Studying
the Effects of Micro-
gravity on Model Parti-
cle/Cell Systems"
- XXXIII. Zalik, Richard:
Auburn University "Rotordynamic Analysis
of a Bearing Tester"
- XXXIV. Ziebarth, John:
The University of
AL in Huntsville "Space Shuttle Main
Engine Numerical Model-
ing Code Modifications
and Analysis"

N89-21727

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

CRYOGENIC FLUID MANAGEMENT IN SPACE

Prepared By:	Basil N. Antar
Academic Rank:	Professor
University and Department:	University of Tennessee Space Institute Engineering Science & Mechanics
NASA/MSFC:	
Laboratory:	Propulsion Laboratory
Division:	Propulsion Systems Division
Branch:	Space Propulsion Branch
MSFC Colleague:	John M. Cramer
Date:	August 10, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

CRYOGENIC FLUID MANAGEMENT

by

Basil N. Antar

Professor, Engineering Science and Mechanics

The University of Tennessee Space Institute

Tullahoma, Tennessee

ABSTRACT

Many future space based vehicles and satellites will require on orbit refuelling procedure. Cryogenic fluid management technology is being developed to assess the requirements of such procedure as well as to aid in the design and development of these vehicles. Cryogenic fluid management technology for this application could be divided into two areas of study, one is concerned with fluid transfer process and the other with cryogenic liquid storage. This division is based upon the needed technology for the development of each area. In the first, the interaction of fluid dynamics with thermodynamics is essential, while in the second only thermodynamic analyses are sufficient to define the problem. In this report we discuss the following specific processes related to the liquid transfer area: tank chilldown and fill; tank pressurization; liquid positioning; and slosh dynamics and control. These specific issues are discussed in relation with the required technology for their development in the low gravity application area. In each process the relevant physics controlling the technology is identified and methods for resolving some of the basic questions are discussed.

INTRODUCTION

Satellites, orbital vehicles, and manned space stations may remain in space permanently in the future. Among the missions planned are the Space-based Transfer Vehicles (STVs), the Space Station, the long life version of the Shuttle Infrared Telescope Facility (SIRTF), the Large Deployable Reflector (LDR), and the Gravity Probe B (GPB) to name a few. All of these missions require cryogenic liquids for fuel or coolant which must be replenished on orbit periodically from tankers which are stationed in orbit. Cryogenic fluid management has become an important component of the design and development of these future vehicles. Thus cryogen transfer and the long term storage of cryogens will be required as an enabling technology for future space missions.

Cryogen fluid management studies began with the advent of the space program due to the use of liquid hydrogen and oxygen for fuel. More than two decades ago a need by NASA developed for controlling propellant position in low-gravity environment with regards to its Apollo program. The Saturn V/S-IVB stage had to be restarted in Earth's orbit after coasting for an extended period in order to place the Apollo in a translunar trajectory. For this end several studies on liquid orientation and cryogen storage in low gravity environment were conducted culminating in an on orbit experiment to gather the necessary data required for the design of the vehicle (Swalley et al., 1966). This was part of an ongoing extensive program to study the effects of low gravity on cryogenic fluid behavior and storage which gave rise to a wide ranging research effort (see for instance Abramson 1966). Needless to say that a vehicle was successfully designed and flown.

With the subsequent advent of the Shuttle program attention was focused more on the storage and the transfer of large amounts of cryogens in terrestrial environment. This was needed to supply the external tank with the necessary liquid fuel. With this program the emphasis was shifted to cryogen management of large quantities without particular attention on cryogen losses and problems concerned with low gravity liquid positioning. However, the low gravity cryogen cooling program necessary for the various satellite operations continued to push the small scale cryogen management technology development by

NASA.

Research on low-gravity cryogenic fluid management continues to be pursued by NASA as part of the ongoing space program. Specifically, two programs dealing with low gravity cryogenic fluid management are being pursued simultaneously, one is directed towards cryogenic cooling applications while the other dealing with the basic scientific questions connected with liquid fuel management applications. Cryogenic cooling was found to be necessary in a variety of space instruments in which NASA continues to play a leading role in developing (see for instance Sherman 1982). On the other hand, there remained unresolved many basic scientific questions with regards to the role of gravity in two-phase flow and heat transfer at cryogenic temperatures. For this objective selected problems dealing with fundamental issues for which it was thought more research was needed were identified and subjected to intensive examination. Major among these issues is liquid slosh in various tank geometries, tank chilldown and fill, and long term cryogen storage. In all of these issues two-phase flows and heat transfer appear to be very important. A major component of this effort is a planned space flight experiment to gather the urgently needed data for that specific technology area (Aydelott 1985). Basically, the experiment consisted of two tanks connected by a transfer line to be placed in the Shuttle cargo bay and in which the working fluid was a cryogen. Several specific issues are to be investigated among which is the receiving tank chilldown and fill, transfer line chilldown, liquid acquisition devices performance and tank pressurization. Recently this experiment has been upgraded into a stand alone space flight experiment in which the working fluid is liquid hydrogen (LH2). However, the technical issues to be investigated remained the same (see Aydelott and Devol, 1987).

The objective of this report is to study four of the major technical issues relevant to low gravity cryogen liquid management and to provide a timely assessment of the present technology status of these issues. Specifically, the issues in question are: tank chilldown and no-vent fill, pressurization system performance, liquid reorientation, and slosh dynamics and control. This study was conducted with application to the Space-based Transport Vehicle liquid propellant on orbit refill in mind.

DISCUSSION

As discussed in the introduction the on-orbit STV refuelling process is an important component of the function of the vehicle. This process requires a thorough understanding of the various aspects of cryogen liquid transfer and storage in space. Cryogen liquid transfer process involves several tasks some of which are: transfer line chilldown, receiving tank chilldown and fill, liquid orientation, and storage vessel drainage. While cryogen storage process is primarily concerned with: pressure control, insulation, and slosh control. Of the tasks enumerated above it was decided to concentrate on the following:

- tank chilldown and no-vent fill,
- pressurization system performance,
- liquid reorientation, and
- slosh dynamics and control.

These issues will be discussed in this section with respect to the important physical processes involved in each and the state of the art level of understanding of the physics of each process. With the understanding that the interest in here is in the STV design, the impact of low gravity environment on these issues must be considered. This requires an assessment of low-gravity testing as well as terrestrial testing.

1. Tank Chilldown and No-Vent Fill

Tank Chill down and fill constitutes a crucial step in any refuelling process since this process will be repeated every time a refill of the STV is accomplished. This process will involve initially of lowering the wall temperature of the empty receiving tank from ambient to a temperature close to the liquid temperature. This is basically accomplished through the introduction of a small charge of liquid into the tank either through a spray or a jet which should contact the tank wall. In this instance heat is rejected from the wall through conduction and is absorbed by the liquid in the form of latent heat of vaporization. This process is repeated until the tank wall reaches the desired temperature. However, due to the liquid evaporation mechanism the tank pressure will rise.

The amount of coolant needed to lower the tank temperature from its ambient value

to a final value can be easily calculated through a thermal balance analysis if the thermodynamic properties of both the coolant and the tank material are known. However, the time needed to accomplish this task is not as easily calculable. The cooling rate is a strong function of the cooling technique employed. The method for calculating this rate varies according to whether a jet is used or a spray. In both cases, however, the cooling rate is predicted by analyzing the hydrodynamic behavior of the cold liquid upon impingement on a hot surface. As an example, the cooling rate will be different for a cold drop impacting a hot surface and for a liquid film moving on a surface. For both cases realistic hydrodynamic-thermal models can be constructed to predict cooling rates once the fluid behavior at and during impact is known. Fluid behavior upon impact can be classified through experimental means. For a drop for instance, it is well known that its shape evolution, and hence the heat transfer rate, when it impacts a hot plate is a function of the Weber number (Bolle and Moureau 1982). The Weber number is the ratio of the drop's kinetic energy to its surface energy due to surface tension. In order to determine the cooling rate it is possible to use the extensive amount of research that already exists for both spray cooling and liquid jet cooling. An excellent review of this subject is given by Bolle and Moureau (1982).

Of course all of the work performed to date on spray and jet cooling has been for terrestrial conditions. It is hard to believe that the hydrodynamics and thermodynamics of this problem is substantially affected by gravity. One area where gravity influence may be felt is in the maximum drop size that can be obtained under low-gravity conditions. It is possible to imagine that drop morphology upon impact will be different for larger drops. However, this problem can be suppressed by imposing a maximum allowable drop size.

There are two deficiency areas in our knowledge that can be immediately identified. One is whether liquid drops disintegration upon impingement on a wall at cryogenic temperatures does obey the established Weber number classification for water? The second is whether this classification is also valid for low-gravity impact environment. Both of these questions can be easily answered with a few simple experiments using either drop towers or parabolic trajectory airplane flights.

Once the tank wall temperature has reached the desired value the filling process of the tank may commence. It is known that the initial amounts of the incoming liquid may vaporize upon entering the tank which will cause the pressure inside the tank to rise. Due to imposed structural constraints the tank pressure can not be allowed to exceed a set value. Thus the tank pressure must be lowered. For the low gravity fill applications it is desirable to achieve the low pressure without venting any of the cryogen. This last condition is imposed to save as much of the fuel as possible since fuel losses cannot be tolerated. The obvious way to achieve no vent fill is to allow the gas inside the tank to condense by further reduction of the tank temperature. It is possible to calculate the liquid mass and degree of supercooling needed to achieve a specific percentage of tank fill under the maximum allowable pressure constraint.

The pressure rise inside tanks of simple geometries (cylindrical, spherical etc.) during a no vent fill process can be readily calculated from the hydrodynamic-thermodynamic governing equations. The solution to these equations may be obtained using computational fluid dynamics (CFD) techniques. In addition to the transient pressure history the temperature of both the fluid and the tank wall and other related field variables could be calculated through these means. Earlier attempts to obtain such solutions via CFD techniques were undertaken by Barakat et al. (1966) and Merte et al. (1970). However, CFD has advanced tremendously in the past 20 years permitting realistic and meaningful solutions to be obtained for the no vent fill process without much trouble.

Basically, the no vent fill process can be modeled by the incompressible mass, momentum and energy conservation equations for the liquid segment and their compressible counterpart for the gas space. Vaporization and condensation on both the solid-gas and liquid-gas interfaces can be accounted for through balances of mass and energy across these interfaces. Also, the liquid influx into the tank can be handled through appropriate boundary conditions. The only problem with this analysis for in gravity applications is the initial position of the liquid-gas interface. In terrestrial environment the interface equilibrium position in a motionless liquid can be assumed to be flat in a plane normal to the gravity vector. However, in zero- or low- gravity environment the gas-liquid interface

is dominated by the surface tension force which is a function of temperature. It can be shown (Reynolds and Satterlee, 1966) that the hydrostatic, stable, equilibrium liquid-gas interface position in an isothermal zero-g environment is one in which the total potential energy is minimized. Thus it is obvious that for an environment such as the no vent fill which is not isothermal an assumption on the initial position of the interface is no trivial matter. Under hydrostatic conditions, the minimization of the surface energy problem reduces to a purely geometric problem. Some calculations have been performed for stable meniscii shapes in spherical tanks under zero-g conditions. However, we can not find any evidence of experimental attempts to verify these calculations.

It is possible to extend the above discussed calculations of Reynolds and Satterlee to the dynamic conditions appropriate to the fill process and thus produce some predictions on shapes and positions of the gas-liquid interface. Once the difficulty connected with the meniscii shapes is resolved then the pressurization problem during no-vent fill may be handled conveniently.

The important physical processes involved in the tank chilldown and fill is clearly the coupling between fluid dynamics and thermodynamics as it relates to cold drop impact on a warm surface and the initial position of the liquid gas interface. We need to understand the drop impact mechanism as well as the flow of liquid in a tank in low gravity. Some low g tests are needed for the drop impact problem but none is needed for the no vent fill problem. Clearly for the drop impact problem the Weber number is important while for the no vent fill the Reynolds number as well as the Bond number are important. The drop impact experiment should be performed with a cryogen liquid in order to fully evaluate the effects of large temperature gradients on this problem. It is immaterial what liquid to use.

2. Pressurization System Performance

The problem of moving the liquid fuel from the storage tank to the receiving tank can be considered as a complimentary problem to the tank fill process. This problem arises whenever a STV fuel tank is needed to be filled on orbit. This process is also of

some concern in cryogenic engineering under terrestrial environment. Under terrestrial conditions Barron (1985) cites three methods that are commonly used to drain cryogenic liquids from the storage vessel. These methods are: (1) self pressurization, (2) external gas pressurization, or (3) pump transfer. It is natural to assume that these same methods are utilized in low gravity applications. The consensus among the low gravity fluid management community has been to employ the pressurization technique, whether self or external, for the on orbit liquid drainage process (see Aydelott and Devol, 1988). Pressurization method of drainage involves creating the pressure differential needed to expel the liquid by injecting gas in the ullage space to increase the gas pressure. In self pressurization, the gas injected is the same as the ullage gas while external pressurization involves introducing high pressure gas from an external source.

From the fluid-thermodynamic point of view the pressurization and the no vent fill processes are analogous with probably the only difference between the two being the tank size. It is possible, then, with simple thermodynamic analyses to calculate the amount of pressurant gas required to induce a specific mass flow rate using heat and mass balances. Such calculation in which the saturation rule is used may be found in Epstein (1965). This technique is known as the lumped system method for calculating pressurization requirements in which only the mean properties of the gas space and tank wall are determined (Barron 1985). However, when more details on the pressurization process are required such as the temperature, composition, velocity or pressure as functions of both space and time, then a distributed analysis is necessary. Such requirement arises whenever a specific control of the process is contemplated. For the distributed analysis again the governing conservation equations of mass, momentum and energy must be solved in both the liquid and the gas. Clark (1965) gives a very good review of the then state of the art for such calculations in terrestrial environment. The similarity between the fill and pressurization processes for the distributed system is obvious in that the necessary calculations are basically CFD calculations which has advanced significantly in the past twenty years.

It is possible to simulate the pressurization and drainage process numerically, provided the initial position of the liquid surface is known and the pressurant gas speed is modest.

The CFD models envisioned for use in these analyses must account for condensation and evaporation across the interface and along the gas-solid boundaries. CFD models can then predict with reasonable accuracy the pressure history and other field variables as well as the position of the liquid in the tank. However, this solution will be appropriate to the specific initial conditions imposed. It is these initial conditions that need to be determined with some confidence.

Again, as in the no vent fill case, in the pressurization process the liquid-gas interface position and shape appear to be an obvious place where gravity may play a role. Thus an analytical model which could predict the interface characteristics in low gravity is essential for distributed system analysis. However, the problem here is not as critical as in the no vent fill process. The position of the inlet nozzle for the pressurant may be open to either the gas or the liquid without affecting the pressurization process itself. On the other hand, the position of the liquid outlet nozzle is very critical in order to allow only liquid to be expelled. Thus it is imperative in this case to be able to predict the liquid position in the storage tank during draining. It is well known from previous work (e.g. Reynolds and Satterlee, 1965) that wetting liquids tend to wet more under low gravity conditions. Hence it is safe to assume that a nozzle placed flush with the tank wall will only allow liquid to be expelled if the gas pressure is not too great so as to break up the liquid surface.

The pressurization process appears to be amenable to moderate CFD modelling effort which needs to be formulated specifically for the liquid expulsion task. However, substantial amount of analytical modelling need to be performed in order to accurately predict the initial conditions, specifically, the liquid surface initial position. The analytical modelling effort must be augmented with simple ground based as well as flight experiments. The effort in this task should be directed towards understanding of the role of surface tension forces, in the absence of the gravity force, in defining the liquid-gas interface.

The pressurization system performance is dependent upon the interaction of thermodynamics with fluid dynamics. This whole technology area can be realistically modeled via CFD analysis. The most important parameter in this problem is the Bond number. It does not appear that a flight experiment is necessary in order to predict the pressurization

process or pressure collapse.

3. Liquid Orientation and Slosh Dynamics and Control

The technical issues related to liquid orientation in low gravity and slosh dynamics are closely related in the sense that both processes are primarily concerned with only fluid dynamics with minimum thermodynamic effects. Consequently, these two issues will be discussed together in here. Of all the technical issues involved in the cryogenic fluid management technology the problem of slosh dynamics and control has received the greatest amount of attention over the years. For this reason there exists a substantial amount of information, both analytical and experimental dealing the low gravity aspects of this problem. The problem of slosh dynamics is of direct importance to space technology especially for liquid propulsion launch vehicles. Such devices have an enormous percentage of their initial weight as fuel and consequently the dynamic forces resulting from the motions of these large liquid masses could be very substantial even beyond the capabilities of the control system to counteract them or the structure to resist them. If the dominant fuel slosh frequencies are close to any of the control system frequencies, an instability of the flight characteristics can result; while if the slosh frequencies are close to the elastic body bending frequencies a large amplitude dynamic response problem may arise.

Due to the criticality of this issue to the launch vehicle design and performance and hence to the space program in general a great amount of research has been devoted specifically to study the problem of slosh dynamics. A review of the available literature at that time is given in Abramson (1965). However, due to the fact that the launch vehicles dispose of most of their fuel in the very first few minutes of launch, the influence of low gravity on this problem was not of primary concern in that body of research. With the subsequent prolific use of satellites and also the expansion of the space program to interplanetary flight a need developed for understanding slosh dynamics and control in low gravity environment. This need resulted in a substantial amount of research which is documented in various books and monographs the most recent and comprehensive of which is given in the book by Myshkis et al. (1987).

Since the original primary goal of the study of slosh dynamics was concerned with the identification of the slosh frequencies for various tank geometries, vibrational environment and fluid fill configurations the problem was resolved through analytical models. In order to solve these models in a straight forward manner specific attention was paid to the linearized form of these models. The linearized system can readily identify the various slosh frequencies without resolving the interaction problem. For the resolution of that latter problem one must turn to the nonlinear form of the models or nowadays to CFD techniques. The only difficulty with the linearized models is the tank shape (i.e. spherical, cylindrical etc.). The more complex the tank geometry is the greater the amount of work needed to determine the slosh frequencies.

The general model for the slosh dynamics problem as well as the fluid orientation problem is constructed by assuming the liquid in the container to be inviscid, irrotational, and incompressible whose velocity field u_i is given by $u_i = u_i(x_i, t)$. This velocity field is governed by the equations of motion given by:

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho} \nabla p - \nabla \Pi, \quad (1)$$

where $p = p(x_i, t)$ is the pressure field in the liquid and Π is the potential function for the body force including gravity. Since the fluid is assumed to be ideal then the velocity field is describable by the potential function, Φ :

$$u_i(x_i, t) = -\nabla \Phi(x_i, t). \quad (2)$$

Which with the conservation of mass equation leads to Laplace's equation for Φ :

$$\nabla^2 \Phi = 0 \quad (3)$$

The shape of the liquid gas interface must satisfy Laplace's condition on the pressure drop across the interface given by:

$$p_o - p = \sigma(k_1 + k_2), \quad (4)$$

where $p_o = \text{const.}$ is the pressure of the gas, while k_1 and k_2 are the principal normal sections of the interface. σ is the surface tension. The model for slosh dynamics must also

satisfy the Dupre-Young condition on the contact line at the triple contact point of the liquid, gas and solid given by:

$$\sigma(\cos\alpha) = \sigma - \sigma_o, \quad (5)$$

where α is the contact angle of the liquid.

In this model for the slosh problem the gas pressure p_o is assumed to be constant and known throughout the gas space and hence no equations are needed for its description. To complete the problem the usual inviscid solid-liquid boundary conditions:

$$u_n = -\partial\Phi/\partial n = 0 \quad (6)$$

must be satisfied on all solid walls. n is the unit outward normal. Also on the liquid-gas interface the condition that the rate of displacement of the interface is equal to the velocity component normal to the interface is imposed:

$$\partial\Phi/\partial n = -\partial N/\partial t, \quad (7)$$

where $N(x_1, x_2, t)$ is function describing the gas liquid interface. Note, since the thermodynamic effects are neglected for this problem, no condensation or evaporation is allowed to take place across the interface.

It should be observed that the governing field equation is linear, in this case it is Laplace's equation, Eq. (3) above. The nonlinearity of the problem in this model is in Laplace's condition, Eq. (4) which describes the interface curvature. The linearized slosh dynamics problem is obtained from the linearization of this condition. The linearized problem is normally cast in the form of an eigenvalue problem in which the eigenvalues determine the slosh frequencies. These frequencies are the slosh frequencies appropriate to the specific tank shape and liquid fill level. These frequencies have been identified for various simple tank shapes. Such an eigenvalue solution is normally considered an analytical solution even though in some cases it requires a moderate amount of numerics.

When condition (4) is not linearized then the problem can not be described as an eigenvalue problem and a more complicated procedure for the solution of the problem

must be adopted. There are basically two ways of solving such nonlinear problems, one is through numerical approximation of the solution and the other through an eigenfunction expansion in terms of the eigenfunctions of the linearized problem. The first method is basically a CFD technique which is fairly common nowadays. In fact there exists a numerical code developed specifically for the slosh dynamics problem available for use, Torrey et al. (1987).

It is clear from examining the governing equations shown above that gravity plays an important role in this problem. In fact the dominant parameters, besides the liquid and gas properties, as they appear in equations (1) - (7) are: gravity (this is implied in the potential function for the body force, Π), the surface tension, σ and the contact angle, α . Thus any solution, whether linear or nonlinear should be in terms of these parameters. The two forces that dominate this problem are clearly the force of gravity and the capillary force whose ratio is the Bond number. Thus it is appropriate to take the Bond number as the single most important parameter in any numerical or experimental simulation. It is clear that under low gravity conditions the capillary forces are dominant. Thus any free surface configuration, as well as the dominant frequency can, in principle, be calculated for the appropriate Bond number and imposed vibrational frequency. Once the shape and the dynamics of the free surface is known for a given set of conditions, then the slosh control problem may be tackled depending on the desired requirements.

The liquid orientation problem, since it requires determining the motion of the bulk of fluid subject to a specific force must be handled through CFD techniques. The formulation of the problem is subject to the same governing equations used for the slosh problem in a slightly modified form. Equation (1) may be modified to include viscous effects, in which case the velocity field can not be written in terms of a potential function. However, equations (4) through (7) above for the interface and boundary conditions must hold. Given a specific external force and the initial position of the liquid in the tank then final position of the liquid can be determined using numerical means such as VOF3D code (Torrey et al. 1987). This problem is very similar to the large amplitude and interaction slosh dynamics problem.

REFERENCES

- Abramson, H. N. 1966 The Dynamic Behavior of Liquids in Moving Containers with Applications to Space Vehicle Technology. NASA SP-106.
- Aydelott, J. C. 1985 Technology Requirements to be Addressed by the NASA Lewis Research Center Cryogenic Fluid Management Facility Program. AIAA Paper AIAA-85-1229, AIAA/SAE/ASME/ASEE 21st Joint Propulsion Conference.
- Aydelott, J. C. and Devol W. 1988 Cryogenic Fluid Management Technology Workshop: Volume-II Roundtable Discussion of Technology Requirements. NASA Conf. Pub. 10009.
- Aydelott, J. C. and Devol W. 1987 Cryogenic Fluid Management Technology Workshop: Volume-I Presentation material and Discussion. NASA conf. Pub. 10001.
- Barakat, H. Z., Merte, Jr. H. and Clark, J. A. 1966, Finite Difference Solutions of Stratification and Pressure Rise in Containers, in *Proceedings of the Conference on Long Term Cryo-Propellant Storage in Space*. NASA Marshall Space Flight Center Report, 145.
- Barron, R. F. 1985, Cryogenic Systems. Cambridge University Press. 379.
- Bolle, L. and Moureau, J. C. 1982, Spray Cooling of Hot Surfaces, in *Multiphase Science and Technology*, G. F. Hewitt, J. M. Delhay and N. Zuber, editors, vol 1, 1.
- Bowles, E. B., Dodge, F. T. and Green, S. T. 1987 Scaling Trade Studies of the Cryogenic Fluid Management Flight Experiment. Southwest Research Institute Final Report SwRI project 06-1427.
- Clark, J. A. 1965, A Review of Pressurization, Stratification, and Interfacial Phenomena, in *Advances in Cryogenic Engineering*, vol 10B, 259.

Epstein, M. 1965, Prediction of Liquid Hydrogen and Oxygen Pressurant Requirements, in *Advances in Cryogenic Engineering*, vol 10B, 303.

Merte, Jr. H., Clark, J. A., Lady, E. R. and Suh, C. C. 1970, Transient Pressure Rise of a Liquid-Vapor System in a Closed Container Under Variable Gravity, in *Heat Transfer 1970*, U. Grigull and E. Hahne, editors, vol VI, B8.8.

Myshkis, A. D., Babitskii, V. G., Kopachevskii, N. D., Slobozhanin, L. A. and Tyuptsov, A. D. 1987, Low-Gravity Fluid Mechanics. Springer Verlag.

Reynolds, W. C. and Satterlee, H. M. 1966, Liquid Propellant Behavior at Low and Zero g, in *The Dynamic Behavior of Liquids in Moving Containers with Applications to Space Vehicle Technology*, H. N. Abramson, editor. NASA SP-106, 387.

Sherman, A. 1982 History, Status and Future Applications of Spaceborne Cryogenic Systems, in *Advances in Cryogenic Engineering*, vol 27, 1007.

Swalley, F. E., W. D. Ward and L. E. Toole 1966, Low Gravity Fluid Behavior and Heat Transfer Results from the S-IVB-203 Flight, in *Proceedings of the Conference on Long Term Cryo-Propellant Storage in Space*. NASA George C. Marshall Space Flight Center. See also "Evaluation of AS-203 Low Gravity Orbital Experiment". Chrysler Corporation Tech. Report HSM-R421-67.

Torrey, M. D., Mjolsness, R. C. and Stein, L. R. 1987, NASA-VOF3D: A Three-Dimensional Computer Program for Incompressible Flows with Free Surfaces. Los Alamos National Laboratory Report LA11009-MS.

N89-21728

1988

NASA/ASEE Summer Faculty Fellowship Program

**Marshall Space Flight Center
The University of Alabama**

**Finite Element Simulation of Thunderstorm
Electrodynamics in the Proximity of the Storm**

Prepared by:	Michael Edward Baginski
Academic Rank:	Assistant Professor
University and Department:	Auburn University Department of Electrical Engineering
NASA/MSFC:	
Laboratory:	Structures and Dynamics
Division:	Earth Science and Application
Branch:	Remote Sensing
MSFC Colleague:	Dr. Hugh Christian Dr. Richard Blakeslee
Date:	September 15, 1988
Contract No:	NGT 01-002-099 The University of Alabama

Finite Element Simulation of Thunderstorm Electrodynamics in the Proximity of the Storm

by

Michael Edward Baginski
Assistant Professor of Electrical Engineering
Auburn University
Auburn, Alabama

Abstract

At the Eighth International Conference on Atmospheric Electricity observations of electric field, Maxwell current density, and air conductivity over thunderstorms were presented by Blakeslee. The measurements were obtained using electric field mills and conductivity probes installed on a U2 aircraft as the aircraft passed approximately directly over an active thunderstorm at an altitude of 18-20 km. Accurate electrical observations of this type are rare and provide important information to those involved in numerically modeling a thunderstorm. A preliminary set of computer simulations based on this data have been conducted and are described in this paper. The simulations show good agreement with measurements and are used to infer the thundercloud's charging current and amount of charge exchanged per flash.

ACKNOWLEDGEMENTS

I would like to thank NASA/ASEE for allowing me to participate in the Summer Faculty Program at Marshall Space Flight Center. The Program was very well coordinated by Mrs. Ernestine Cothran of NASA and Dr. Michael Freeman of UA. Dr. Hugh Christian, Dr. Richard Blakeslee and Dr. James Arnold were especially helpful in obtaining the necessary computer software and technical materials required for the research.

This summer's research endeavors have lead to the submittal of two manuscripts (submitted to The Journal of Geophysical Research-*special issue* and Geophysical Research Letters), neither of which would have ever come about had it not been for my day to day interactions with Dr. Hugh Christian and Dr. Richard Blakeslee. My appreciation for the personal interest both of these scientists have taken in the research can not be stated strongly enough. I am looking forward to future research in the area of Atmospheric Electricity with Dr. Christian and Dr. Blakeslee

Introduction

The focus of this research is to numerically characterize the electrodynamic behavior above a thunderstorm at altitudes of 18-26 km and lateral distances out to 10 km. This range is selected in light of the recently available electric field and Maxwell current data obtained from U2 flights over an active thunderstorm (presented by Blakeslee *et al.*, at the International Conference on Atmospheric Electricity, Uppsala, Sweden). By comparing simulated behavior to measurements a more complete understanding of the thunderstorm's electrical behavior results, especially concerning the storms electrical sources and its possible influence on the global circuit.

This investigation is based on a finite element solution of the governing differential equations (the Maxwell Equations) and includes the thundercloud charging current and the charge rearrangement following intra cloud lightning [Baginski *et al.*, 1988]. A complete description of the computer code used and its operation is given by Baginski (1987). The scope of the present investigation does not include the atmosphere's response to the high current lightning transient (i.e., propagating electromagnetic energy induced by the lightning return stroke's current) but addresses the electrodynamics following the lightning column's cessation.

Historical Perspective

Since the 1950's several electrical models describing the interaction of thunderstorms with the atmosphere have been published. Holzer and Saxon (1952) have assumed concentrated charges in a dipole configuration with spatially varying conductivities to obtain temporally invariant field patterns in the lower atmosphere and ionosphere. The lightning return stroke, however, generates transients in the electric field pattern known as "field changes" [Uman, 1969]. Early workers attributed this temporal recovery to recharging within the thundercloud. Tamura (1955) is credited as the first to note that the surrounding atmosphere is also involved. He defined solutions based on the conservative field assumption (i.e., $\nabla \times \mathbf{E} = 0$, $\mathbf{E} \sim \exp(-t/T)$, $T = \epsilon/\sigma$)

that depend on the conductivity at the point of observation. Kasmir (1959) constructed the first dynamic model of the thundercloud system using resistors, a capacitor, and a spark gap. His model connected a current generator, a capacitor, and a resistor in parallel to model the cloud ionospheric connections with the path to earth replaced by a resistor. More dynamic models began to follow. Anderson and Freier (1969) incorporated dynamic changes in the dipole structure with spatially varying conductivities. However, Anderson and Freier omit the total set of Maxwell's equations and a dynamic forcing current in their modeling--only the quasi-static relaxation is included. Additional transient solutions were developed based on the "monopole" model of C. T. R. Wilson (1916) by Illingworth (1972), Park and Dejnakintra (1973), Greifinger and Greifinger (1976), and Holzworth and Chiu (1982). To date, there have been many more thunderstorm models presented, several of which solve the self consistent set of Maxwell's equations (Faraday's law of induction and Ampere's circuital law) with a high degree a numerical resolution. The errors inherent to numerical solutions of the transient event are now generally reducible to negligible levels; this advance is due mainly to the advent of the super computer (computational intensive codes are no longer a problem).

Since numerical models will (generally) solve the governing set of equations correctly for the specified input parameters, the focus is therefore to determine accurately the input conditions (e.g., "stiff" sources, boundary conditions, and conductivity) pertinent to the thundercloud's description. Thundercloud information of this type is based on experimental data; a complete description of which is unfortunately not presently available. The measurements presented by Blakeslee *et al.* (1988) provide a significant contribution to the existing data pool available and serve as the basis of the modeling described here.

Measurements

Vertical electric field and conductivity measurements were obtained using U2 borne field mls and conductivity probes above active thunderstorms. The plane flew above the thunderstorm at an altitude of approximately 18-20 km at a speed of ~ 200 m/s. Measurements were

obtained as the plane passed directly over the storm with temporal resolutions of ~ 100 msec. A description of the experiment is given by Blakeslee *et al.* (this issue).

Excerpts of the measured data used for comparison here (Figure 1) represent typical schema of the transient signatures observed. It is the primary intent of this modeling effort to depict the relative magnitudes and temporal signatures observed, with the possibility of predicting the general electrical environment about the storm systems instrumented. The thunderstorms interaction with the Global Electric Circuit is a logical extension of the modeling and will be addressed in future work.

Modeling of the Thunderstorm

The thunderstorms electrical activity is sustained by a constant current generator that exists between upper and lower charge centers (6 and 10 km respectively, upper center positive). Intra cloud lightning, resulting from the accumulation of the generator charge, occurs at time intervals determined primarily by the charging current and amount of charge exchanged per flash. For the purposes of modeling, the effects of the constant current generator will be analyzed separately from that of the charge perturbation associated with intra cloud lightning. The resulting steady state and transient solutions will then be superimposed to determine the total electrical response of interest [Baginski, 1987].

Several implicit assumptions should be noted. The net amount of thundercloud charging caused by effects of charging current and discharging the thundercloud via intra cloud lightning is assumed to be zero (following the initial charge accumulation). This requires that the average amount of charge exchanged via intra cloud lightning equals that of the storm's charging generator. In the modeling a one ampere charging current is used. Charge accumulation at the charge centers continues until a breakdown field strength is obtained [Uman, 1969]. The resulting steady state field mapping at the time breakdown is reached is the steady state component of the relative field signatures of the solutions. Specific details of the spatial and temporal distribution of the thundercloud's charge used in the modeling are given in the following sections.

The computer code used in this modeling is based on an adaptation of a previously developed code that identified middle atmospheric and ionospheric lightning induced signatures [Baginski *et al.*, 1988]. The errors associated with spatial and temporal resolution of the earlier code's results were reduced to minimal levels. Since the simulations of interest to this study are at a significantly closer range and require temporal resolutions ~ 100 msec (instrumental limit), the spatial and temporal discretization used previously will be used here as well.

Charging Mechanisms

As previously described, the thunderstorm's electrical activity is sustained by charge separation which induces a net positive upper and negative lower charge center in a dipole configuration. The height of the charge centers is somewhat affected by seasonal and geographic effects. Heights of 10 km for the upper and 6 km for the lower charge center are not unreasonable [Chalmers, 1967] and will be used for the model [Baginski, 1987]. The current resulting from an intra cloud flash is responsible for the charge perturbation [Uman, 1969] simulated. The rate of deposition of lightning current is proportional to the time derivative of the charge perturbation [Uman, 1969]. Therefore, the total charge deposited at time $<t>$ may be expressed as the integral of the lightning current in time:

$$Q_f(t) = \int_0^t I_{ic}(\tau) d\tau \quad (1)$$

where $I_{ic}(\tau)$ = intra cloud lightning current
 $Q_f(t)$ = total displaced charge

Sunde's (1968) lightning return stroke model is selected for this study. Sunde's model consists of two exponential terms and is relatively simple compared to some [Uman, 1969], but for the time frames of interest in this study ($t > 100$ msec), it includes the necessary temporal information required to predict the late time transient electromagnetic behavior [Sunde, 1968]. The intra cloud charge

perturbation may be expressed in terms of the temporal behavior of this current, as follows:

$$i(t) = I0(\exp(-at) - \exp(-bt)) \quad (2)$$

where $i(t)$ = return stroke current (Sunde's model)

$$a = 10^4 \text{ seconds}^{-1}$$

$$b = 0.5 \times 10^6$$

$I0$ = proportional to amount of charge displaced during return stroke

The temporal structure of the forced charge generator is given as:

$$d Q_f(t)/dt = I0(\exp(-at) - \exp(-bt)) \quad (3)$$

A cylindrical coordinate system is used in the model with symmetry assumed about the vertical (z) axis (Figure 2). The spatial structure of the deposited charge (transient and steady state) is given by a modified spherical Gaussian profile:

$$\pm \partial \rho_f(r,z,t)/\partial t = \pm (\partial(Q_f(t)/\partial t \pm \alpha)(f(r,z)) \quad (4)$$

$$f(r,z) = (\exp(-R/(2\lambda)))/(2\pi\lambda)^{1.5}$$

where λ = variance ($\lambda = 4000 \text{ m}^2$ for simulations)

$$R = r^2 + (z - z')^2$$

z' = altitude of charge perturbation (6 or 10 km)

α = steady state charging current
($\pm 1 \text{ A}$ used in the modeling)

The spatial distribution of the charge perturbation does not noticeably effect the electric field and Maxwell current density signatures far from its interior [Baginski *et al.*, 1988]; since transient phenomenology exterior to the cloud is of interest here a certain degree of freedom exists in the specification of the distribution.

Geometry of the Region

The region selected (Figure 2) is contained within a perfectly conducting right circular cylinder with a radius of 80 km and height of 110 km. The earth's surface is modeled electrically as a perfect conductor (lower plate). Typical values of 10^{-3} to 10^{-2} mhos/meter [Volland, 1984] are given for the earth's conductivity while 10^{-14} to 10^{-13} mhos/meter is the usual range of the adjacent atmosphere's conductivity. This is a difference of more than 11 orders of magnitude, making the earth's surface appear (electrically) as a perfect conductor with respect to the atmosphere. The simulations of interest were found to be insensitive to increases in either the vertical (110 km) or the radial (80 km) limits. The Hall and Pederson components of the conductivity (present above ~ 70 km) are neglected in the formulation [Baginski *et al.*, 1988].

The Maxwell Equations

From the Maxwell Equations a single equation is developed in which the electric field is dependent on the source charge and current densities as follows [Holzworth and Chiu, 1982]:

$$\nabla \times \nabla \times \mathbf{E} = -\mu_0 (\sigma \partial^2 \mathbf{E} / \partial t + \partial \mathbf{J} / \partial t + \epsilon_0 \partial^2 \mathbf{E} / \partial t^2) \quad (5)$$

$$\nabla \rho / \epsilon_0 = \nabla^2 \mathbf{E} - \mu_0 (\sigma \partial \mathbf{E} / \partial t + \partial \mathbf{J}_s / \partial t + \epsilon_0 \partial^2 \mathbf{E} / \partial t^2) \quad (6)$$

where \mathbf{J}_s = source current density associated with intra cloud stroke current, is neglected in the simulations

ρ = charge density

The continuity equation is derived by taking the divergence of the Maxwell current density:

$$0 = \nabla \cdot \nabla \times H = \nabla \cdot (\sigma E + \epsilon_0 \partial E / \partial t + J_s) \quad (7)$$

$$0 = \rho \sigma / \epsilon_0 + \nabla \sigma \cdot E + \partial \rho / \partial t + G_s \quad (8)$$

where $\nabla \cdot J_s = G_s$ = source of charge perturbation
(deposition of intra cloud lightning current)

Equations 6 and 8 describe the electrodynamic response of the atmosphere to the assumed thundercloud charge and current configuration. A charge perturbation of 1 C and steady state thundercloud charging current of 1 A are used in the modeling. Intra cloud lightning is simulated by superimposing the transient signatures resulting from total charge perturbations of + 1 C at $z' = 6$ km and - 1 C at $z' = 10$ km. The conductivity profile used in the previously referenced high altitude simulations [Baginski *et al.* 1988] is slightly modified here to depict the observed conductivity [Blakeslee *et al.*, 1988].

Results

Transient Signatures

Figure 3 identifies the vertical electric field signatures resulting from a positive charge perturbation at 6 km (modeled cloud to ground lightning) and charge perturbations at 6 km and 10 km with the upper charge sensed negative (intra cloud lightning). The magnitude of the waveforms decrease as radial distance is increased with the temporal structure showing only slight changes. This is expected considering previous numerical studies [Baginski, 1987] and observations [Blakeslee *et al.*, 1988]. The simulated intra cloud lightning waveforms are constructed by superimposing the results of perturbations at 6 and 10 km with the perturbation at 10 km sensed negative.

The corresponding vertical Maxwell current density transients are shown in Figure 4. Following the onset of the transient there is an extremely rapid temporal decay in magnitude out to ~ 1 msec followed by a much slower decay [Hale and Baginski, 1987]. The temporal

characteristics of these signatures are very similar for the respective perturbations with magnitude decreasing as radial distance is increased.

Simulations of U2 Measurements

In order to accurately simulate the observed electric field and Maxwell current density patterns the assumed thundercloud charging current and intra cloud charge perturbation are scaled appropriately. The measured electric field and Maxwell current density waveforms shown in Figure 1 are assumed to occur as the plane passes approximately above the thunderstorm [Blakeslee *et al.*, 1988]. There is some uncertainty as to the plane's exact location relative to the thunderstorm and therefore reasonable estimates of the scaling parameters must be made.

After a series of simulations were conducted with a wide range of scaling variance the following parameters were selected:

Cloud charging current ~ 5 Amperes

Amount of charge exchanged during

intra cloud lightning ~ 10 -20 Coulomb

Altitude of observation = 18 km

Time delay prior to 1st intra cloud lightning = 3 seconds

Time delay prior to 2nd intra cloud lightning = 15 seconds

The resulting scaled simulations are shown in Figure 5. There is good overall agreement between the scaled simulations and measurements. The major differences are in the amount of "noise" associated with the measured data. This may be due in part to concurrent thundercloud discharges many orders of magnitude less than the primary discharge. In addition, the simulated Maxwell current density waveforms do not contain large negative transients with temporal durations of ~ 500 msec of as shown in the measurements in Figure 1. This may partially be a result of an actual flash consisting of multiple strokes and partially an artifact arising from insufficiently resolving $\epsilon_0 \partial E / \partial t$ during lightning discharges in the derivation of the Maxwell current density from E.

Conclusions

The results of the study add credence to the belief that accurate computer simulations of the atmosphere's response to thunderstorms are possible. A comparison of the (unscaled) simulated and measured Maxwell current densities and electric fields showed good agreement temporally. Agreement in magnitude was attained by incorporating scaling factors in the simulated results. Since the simulation's governing set of differential equations are linear (the Maxwell Equations), the thunderstorms charging current and net amount of charge exchanged during intra cloud or cloud to ground lightning may be inferred using magnitude scaling.

Future research in electrically modeling a thunderstorm will benefit from a significantly larger data base. Several thunderstorm campaigns that would provide this type of information are being considered for the early 1990's.

REFERENCES

- Anderson, F.J. and G.D. Freier, Interaction of the thunderstorm with a conducting atmosphere, *J. Geophys. Res.*, 74, 5390, 1969.
- Baginski, M.E., "Finite element simulation of the atmosphere's electromagnetic response to charge perturbations associated with lightning," Ph.D. thesis, Pennsylvania State University, 1987.
- Baginski, M.E., L.C. Hale, and J.J. Olivero, "Lightning-Related Fields in the Ionosphere," *Geophys. Res. Let.*, August, 1988
- Blakeslee, R.B., H.J. Christian and B. Vonnegut, "Electrical measurements over active thunderstorms," *Proceedings of the 8th International Conference on Atmospheric Electricity*, 1998
- Chalmers, J.A., *Atmospheric Electricity*, 2nd ed., Pergamon, New York, 1967.
- Greifinger, C., and P. Greifinger, "Transient ULF electric and magnetic fields following a lightning discharge," *J. Geophys. Res.*, 81, 2237, 1976.
- Gish, O.H., and R.G. Wait, "Thunderstorms and the earth's general electrification," *J. Geophys. Res.*, 55, 473, 1950.
- Hale, L.C., and M.E. Baginski, "Current to the ionosphere following a lightning stroke," *Nature*, 329, 814, 1987
- Holzer, R.E. and D.S. Saxon, "Distribution of electrical conduction currents in the vicinity of Thunderstorms," *J. Geophys. Res.*, 57, 207, 1952.

Holzworth R.R., M.C. Kelly, C.L. Siefring, L.C. Hale, and J.D. Mitchell, "Electrical measurements in the atmosphere and ionosphere over an active thunderstorm: 2. Electric field and conductivity," J. Geophys. Res.,

Kasemir, H.W., "The thunderstorm as a generator in the global electric circuit (in German), Z. Geophys., 25, 33, 1959.

Krider, E.P., and J.A. Musser, "Maxwell currents under a thunderstorm," J. Geophys. Res., 11, 171-176, 1982.

Illingworth, A. J., "Electric field recovery after lightning as the response of a conducting atmosphere to a field change," "Quart. J. Roy. Meteorol. Soc., 98, 604, 1972.

Nisbet, J.S., "A dynamic model of thundercloud electric fields," J. Atmos Sci., 40, 2855, 1983.

Park, C.G., and M. Dejnakintra, "Penetration of thundercloud electric fields into the ionosphere and magnetosphere: 1. Middle and subauroral latitudes," J. Geophys. Res., 78, 6623-6633, 1973.

Sunde, E.D., Earth Conduction Effects in Transmission Systems, Dover, New York, 1968.

Tamura, Y., "An analysis of the electric field after lightning discharges," Geophys. Res. Pap., 42, 190-200, 1955.

Uman, M.A., Lightning, McGraw-Hill, New York, 1969.

Volland, H., Atmospheric Electrodynamics, Chemistry in Space, vol. II, Springer-Verlag, Berlin, Germany, 1984.

Wilson, C. T. R., "Some thundercloud problems," J. Franklin Inst., 208, 1916

Figure Captions

1. Measurements made during U2 flights over an active thunderstorm presented by Blakeslee at the International Symposium on Atmospheric Electricity. The vertical electric field, Maxwell current density, conduction current density and displacement current density are shown in from top to bottom respectively
2. Geometry of model simulation
3. Simulated vertical electric field transients at an altitude of 18 km for intra cloud lightning and a charge perturbation centered at 6 km. A total charge of 1 coulomb is exchanged.
4. Simulated vertical Maxwell current density transients at an altitude of 18 km for intra cloud lightning and a charge perturbation centered at 6 km. A total charge of 1 coulomb is exchanged.
5. Simulated vertical electric field and Maxwell current density waveforms using scaling factors at an altitude of 18 km for intra cloud lightning. A total charge of 1 coulomb is exchanged with cloud charging current ~ 5 Amperes, amount of charge exchanged during intra cloud lightning ~ 10 -20 coulombs.

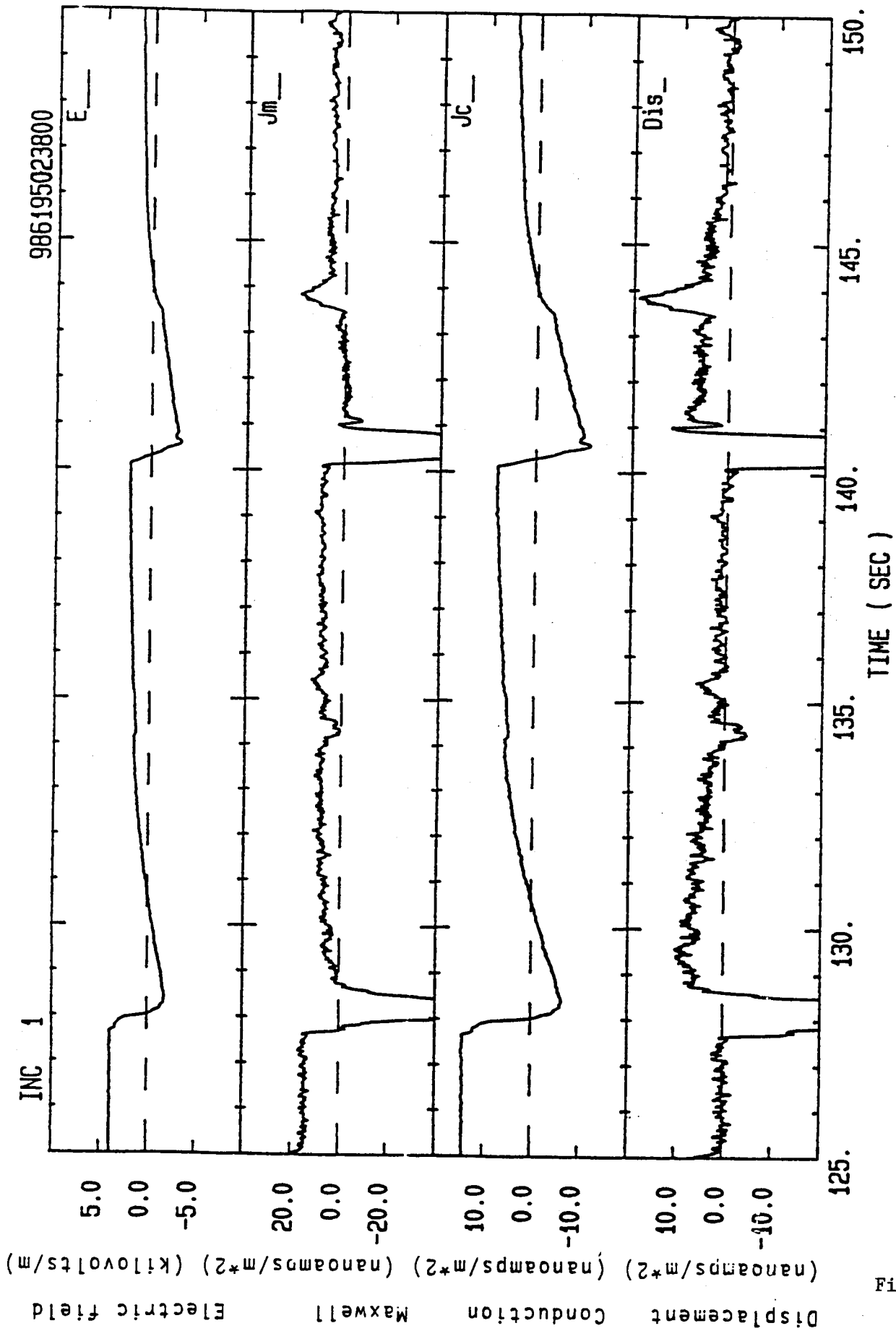


Figure 1
II-13

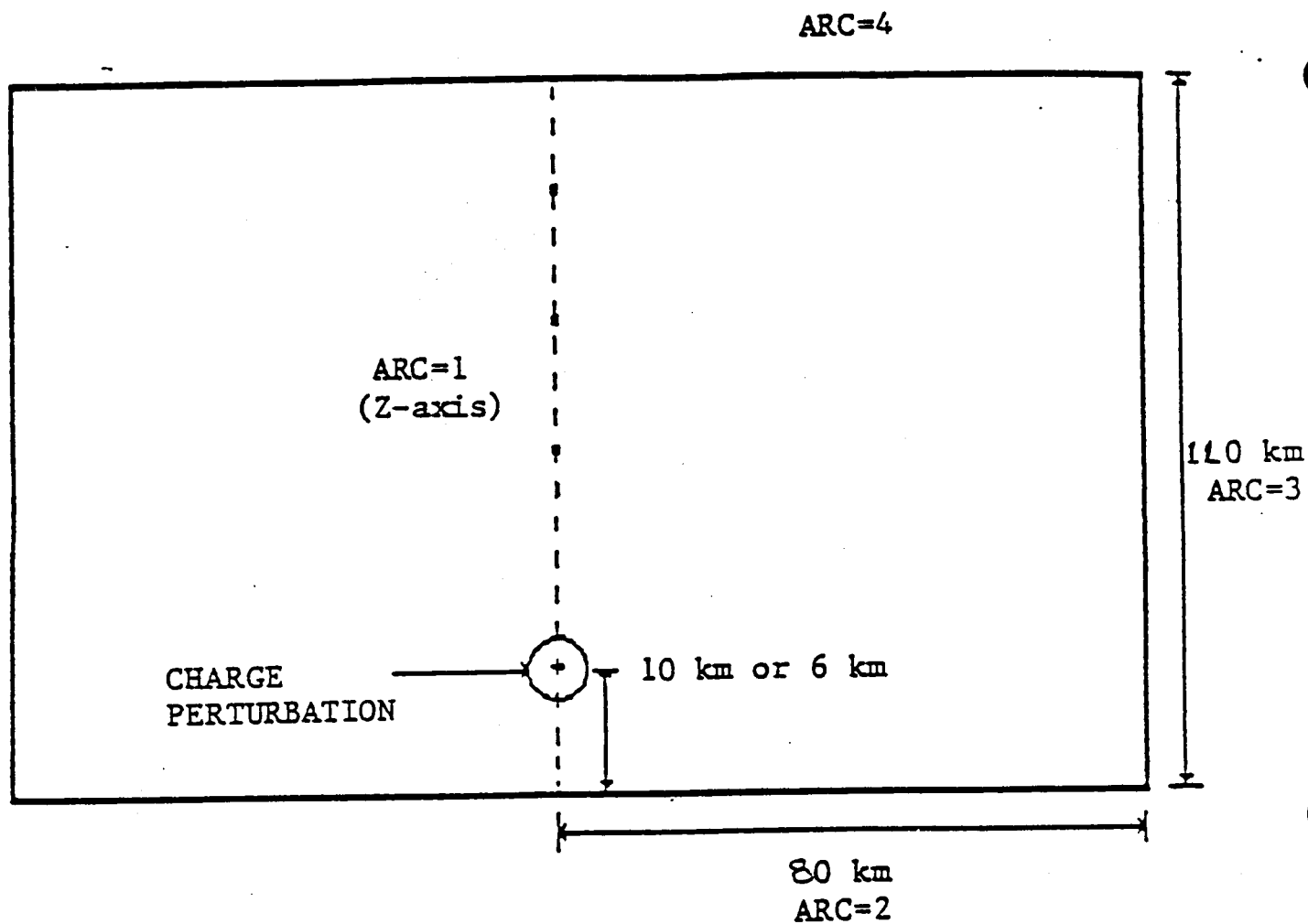


Figure 2.

ELECTRIC FIELD AT Z = 18 K

INTRA CLOUD DISCHARGE OF 1 C

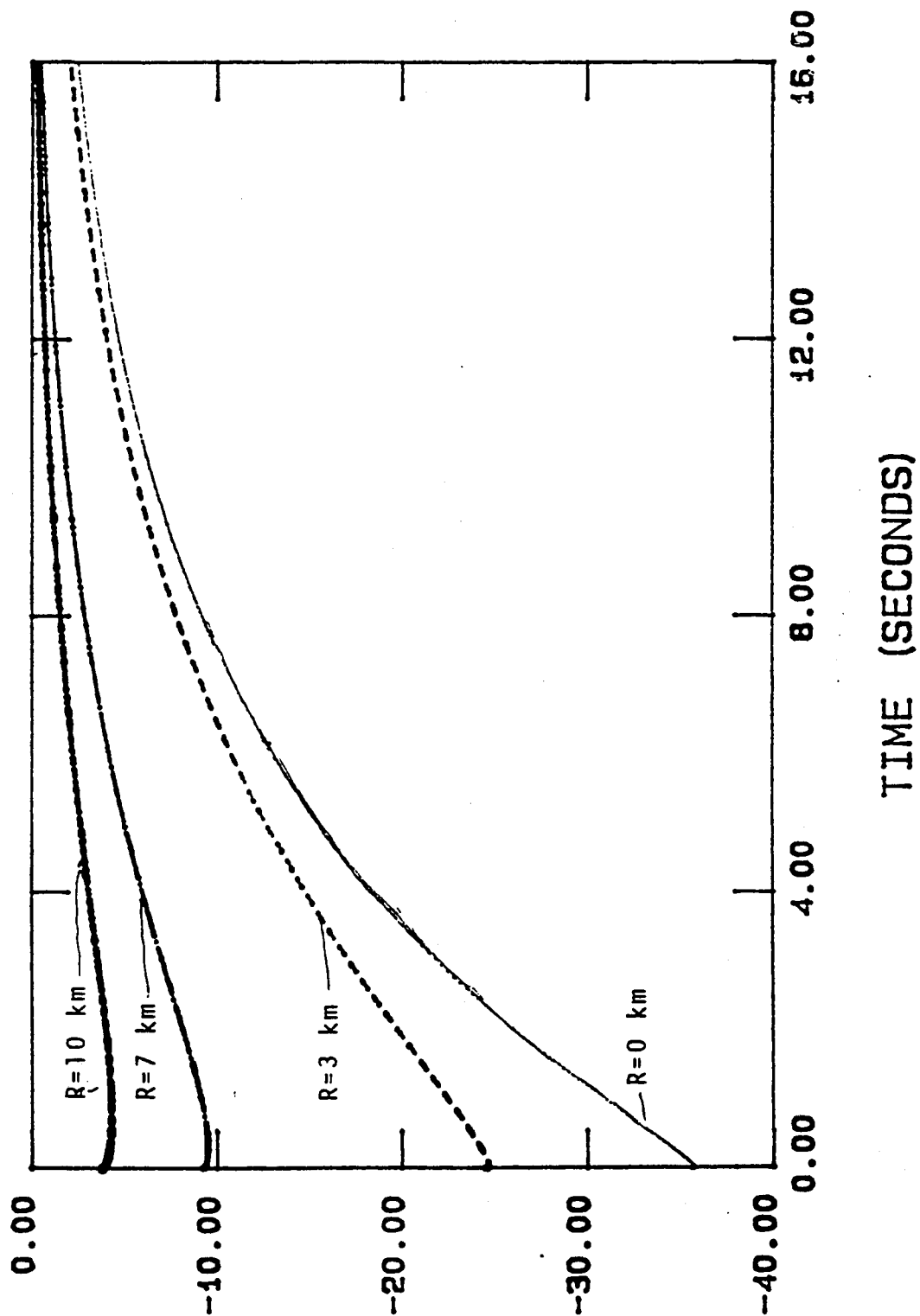


Figure 3a

ELECTRIC FIELD (V/m)

ELECTRIC FIELD AT Z=18 KM

CHARGE PER. OF 1 C AT 6 KM

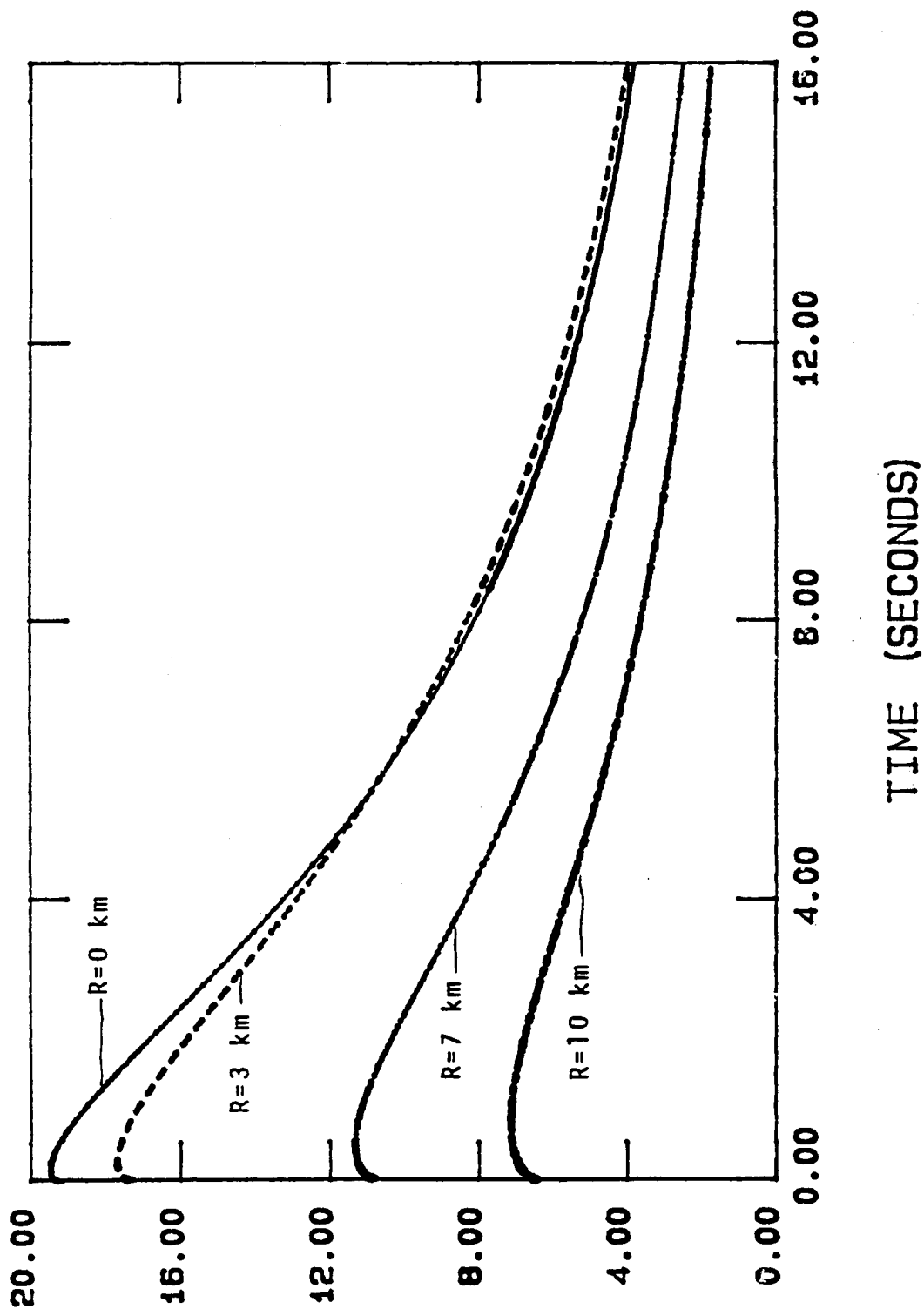


Figure 3b

ELECTRIC FIELD (V/m)

TRANSIENT MAXWELL CURRENT

INTRA CLOUD LIGHTNING. MAG(Jm) AT Z=18 KM

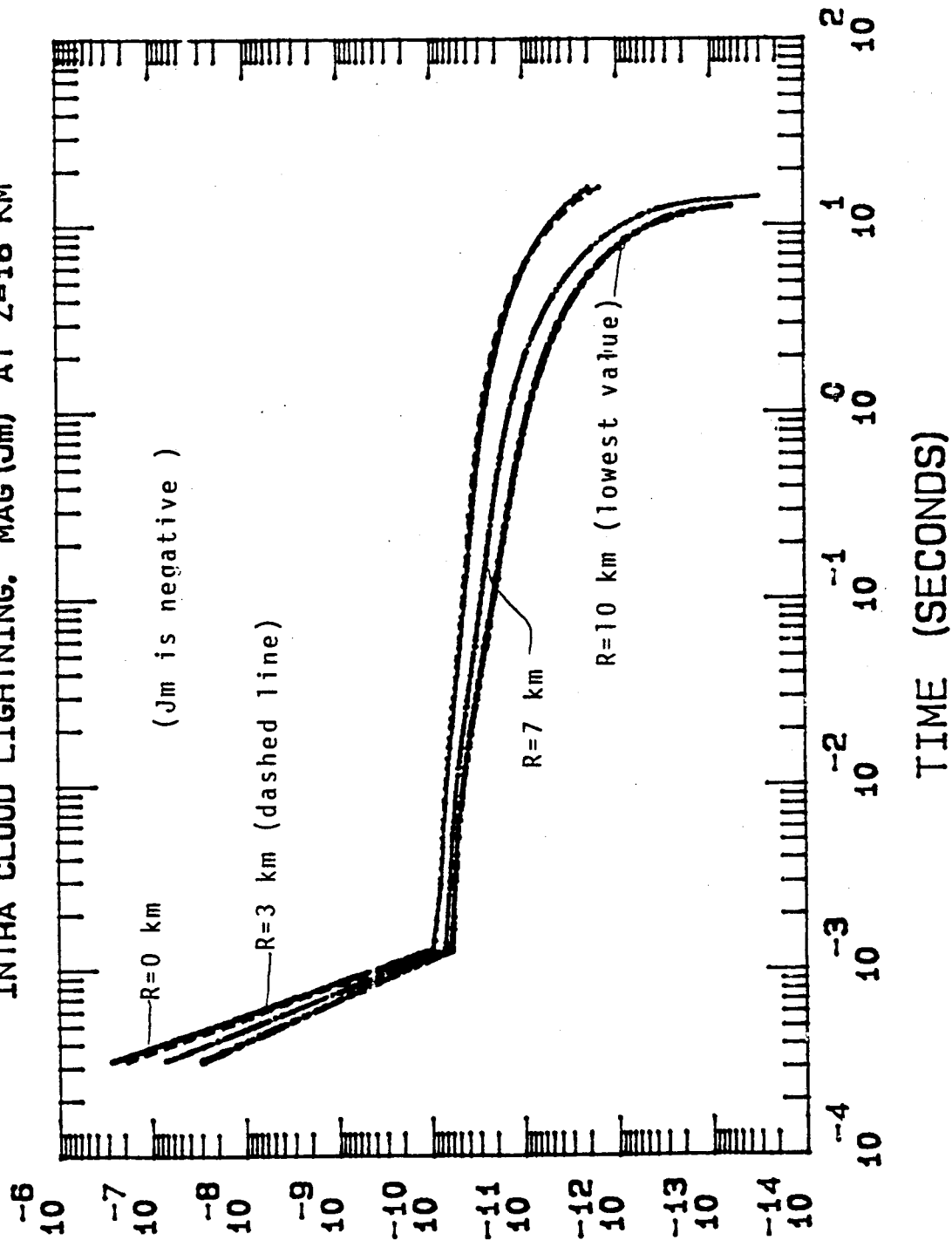


Figure 4a

AMPERES/m²

TRANSIENT MAXWELL CURRENT

CHARGE PER. AT 6 KM. JM AT Z=18 KM

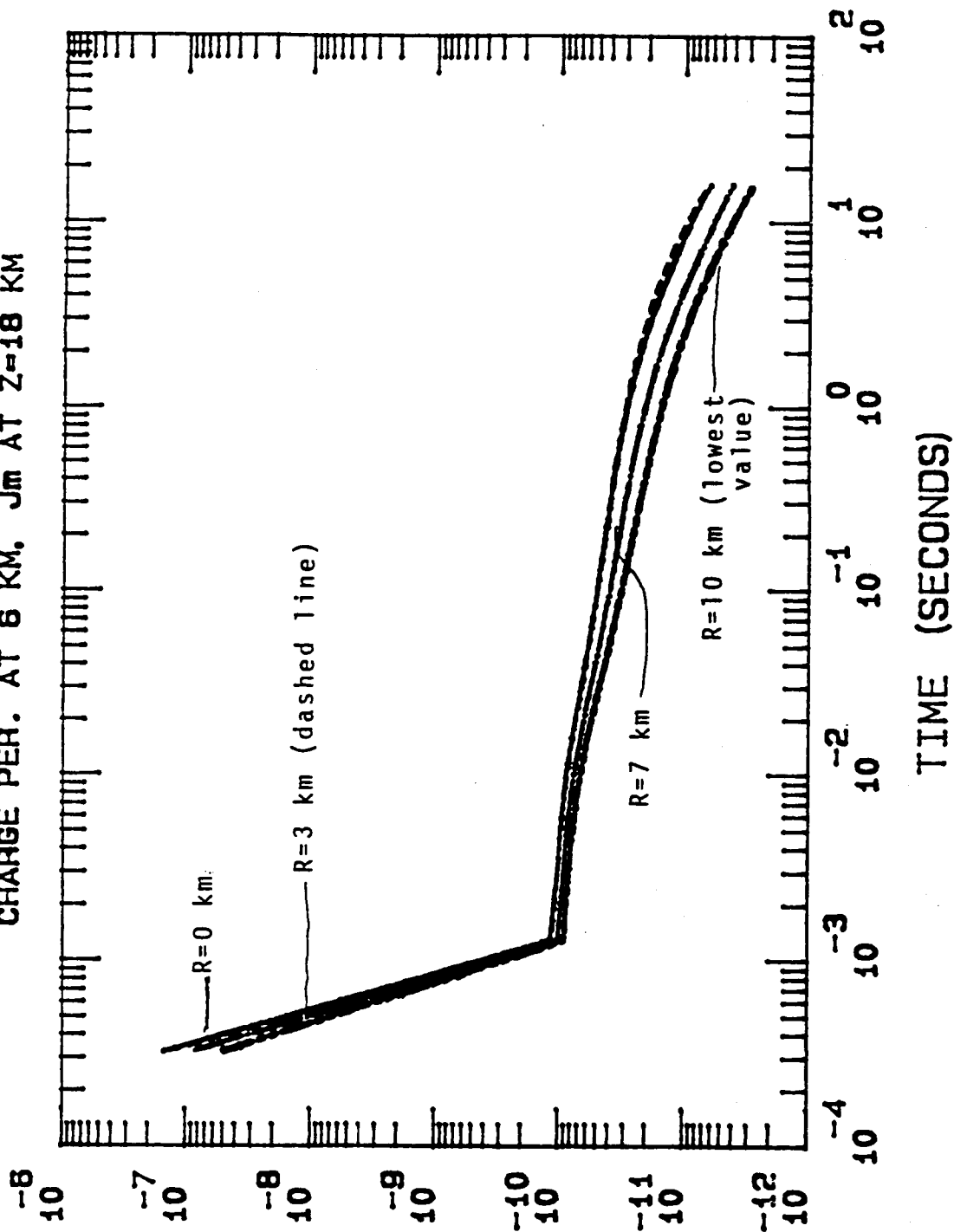


Figure 4B

AMPERES/m²

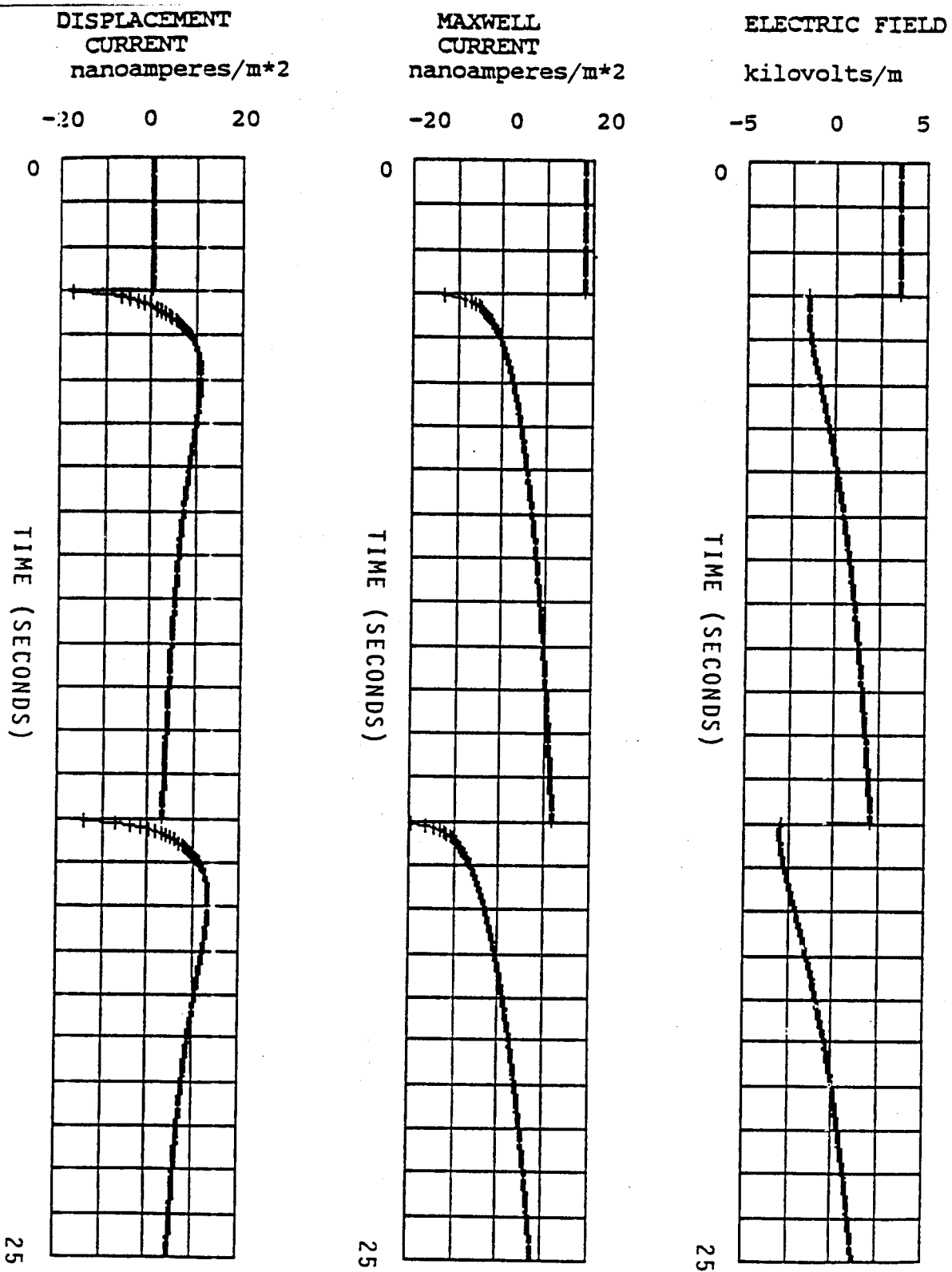


Figure 5

N89-21729

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

REFINING, REVISING, AUGMENTING, COMPILING AND DEVELOPING
COMPUTER ASSISTED INSTRUCTION K-12 AEROSPACE MATERIALS
FOR IMPLEMENTATION IN
NASA SPACELINK ELECTRONIC INFORMATION SYSTEM

Prepared By:	Jean A. Blake
Academic Rank:	Professor
University and Department	Alabama A & M University Mathematics
NASA/MSFC	
Division:	Public Affairs Office
Branch:	Public Services & Education
NASA Colleague:	William E. Anderson
Date:	August 1, 1988
Contract No:	NGT-01-002-099 The University of Alabama

ABSTRACT

"NASA Spacelink is an electronic information service operated by the Marshall Space Flight Center. It contains extensive NASA news and educational resources that can be accessed by anyone with a computer and modem". (4)

Spacelink provides updates and information on:

1. Current NASA News
2. Aeronautics
3. Space Exploration: Before the Shuttle
4. Space Exploration: The Shuttle and Beyond
5. NASA Installations
6. NASA Educational Services
7. Materials for Classroom Use
8. Space Program Spinoffs

I was privileged to participate in the development of Spacelink during the periods of its gestation, birth and infancy. In addition to compiling and developing more material for implementation in Spacelink (including lesson plans and activities for Grades K-12), Summer 1988 was spent refining, revising, and augmenting the material prepared during the previous summer.

Material for the above was extracted from existing NASA publications on aerospace activities as well as from materials developed by other NASA activities (including NASA Educational Workshop for Elementary School Teachers (NEWEST) held at Marshall Space Flight Center, Huntsville, Alabama).

ACKNOWLEDGEMENT

My deepest gratitude is hereby extended to the NASA/ASEE Summer Faculty Fellowship Program and its directors for the very rewarding experience afforded me this summer. I pay special tribute to Dr. C. R. Chappell, Mrs. Ernestine Cothran and Dr. Mike Freeman, the directors of this program whose effervescence permeated the weekly seminars as they gave their support, and guidance. I especially thank my NASA colleague, Bill Anderson, who afforded me the privilege of working on material for Spacelink, and lent his support and encouragement from day to day. The staff in Public Affairs needs special mention because each in a small or large manner contributed to enriching my summer's experience. Special recognition is due Jim Sahli who kindly loaned me his office for the entire 10-week period. A million thanks to one and all!

1. Introduction

The National Aeronautics and Space Administration offers educators a wide range of educational services including speakers, publications, audiovisual materials, software, advanced educational technology, curriculum assistance, electronic communications, workshops, in-school satellite programs, student programs and training opportunities. One of the latest development is the educational service called Spacelink.

"NASA Spacelink runs on a Data General MV-7800 super-minicomputer located at the NASA George C. Marshall Space Flight Center in Huntsville, Alabama. NASA Spacelink software was developed and donated to NASA by the Data General Corporation of Westboro, Massachusetts. The system can communicate with eight callers simultaneously. NASA Spacelink is a dynamic system that will change and expand daily. It was made available to the public in January, 1988.

Initial funding for NASA Spacelink was provided by the Educational Affairs Division at NASA Headquarters. The NASA Spacelink data base is maintained by the Public Services and Education Branch of the Marshall Space Flight Center Public Affairs Office. Operational support is provided by the Information Systems Office at the Marshall Center. Information on NASA scientific projects and educational programs is provided to NASA Spacelink by education specialists at NASA Headquarters and the NASA field centers.

While NASA understands that people from a wide variety of backgrounds will use NASA Spacelink, the system is specifically designed for teachers. The data base is arranged to provide easy access to current and historical information on NASA aeronautics and space research. Also included are suggested classroom activities that incorporate information on NASA projects to teach a number of scientific principles. Unlike bulletin board systems, NASA Spacelink does not provide for interaction between callers. However it does allow teachers and other callers to leave questions and comments for NASA which may be answered by regular mail". (4)

2. The Process

The material, compiled on the word processor, was first loaded into a "working" Spacelink program and thence into the "real" Spacelink program. The material covers the following areas: Aeronautics, Space Exploration: Before the Shuttle, Space Exploration: The Shuttle and Beyond, NASA Installations, NASA Educational Services, Materials for Classroom Use, and Space Program Spinoffs.

Materials for Classroom Use include:

1. Living In Space Activities
2. Space Science Activities
3. Astronomy Information
4. Careers in Aerospace

The lesson plans and activities for living in space cover:

Food
Clothing
Communication
Health
Housing
Working, and information on
Space Station Research and Design.

Other space science lesson plans and activities cover:

Astronauts
Atmosphere
Magnetosphere
Aeronautics
Rockets
Technological Advances
Unmanned Earth Satellites
Unmanned Solar System Exploration
Man in Space
Projections
Solar Cells
Miscellaneous Activities.

The Appendix contains a partial listing of the areas and topics covered.

APPENDIX

A Listing of some of the areas and topics from which documents were prepared for Spacelink, 1988.

Areas

Art
Astronomy
Biology
Chemistry
Communication
Earth Science
Engineering
Fine Arts
Geography
Geology
Health
Home Economics
Language
Life Science
Mathematics
Nutrition
Physical Science
Physics
Political Science
Science
Social Science
Social Studies

Topics

Accomplishments and Benefits of the Space Program
Application Procedures for Employment with Marshall
Space Flight Center
Artificial Intelligence
Atmospheric Flight Research Design and Testing
Bernoullis Law, Airplane Design
Breathing Volume
Career Opportunities in Aerospace Technology
Crystal Growth
Earth Observation
Exercise and Pulse Rate
Exploration of Earth Resources from Space
Flight Design
Flight Research and Exploration

Group Dynamics
Helicopters
Image Interpreting
Imaging Systems
Launch Systems
Launch Vehicles
Living and Working in Space
Living Systems
Lunar Features
Lunar Prospecting
Lunar Science
Magnetic Fields
Manned/Unmanned Space Flight
Map Making
Naming an Orbiter
Navigation
Payload Packaging
Photography from the Air and Space
Planetary Science
Reaction Time
Rocketry and Propulsion
Satellite Communication
Science and Society
Shuttle Layout
Solar Cells
Solar Energy
Solar System
Solar System Exploration
Solar System Research
Space-Age Robotics
Spacecraft Thermal Control
Spacecraft Power Systems
Space Flight
Space Food
Space Poetry
Space Station
Space Telescope
Space Travel
Space Words
Stellar Astronomy
Sunspots
Task Performance
The Astronaut's Hall of Fame
The Nature of Stars
Time in Space
Toys in Space

CONCLUSIONS and RECOMMENDATIONS

Every university professor will underscore the fact that work at the college level is made easier when entering college students are academically prepared for college. Everyone also knows that academic foundation is very important. The elementary school and high school experience should be rewarding and well laid. I, therefore, count it a great privilege to contribute in this small manner to helping elementary and secondary teachers through the medium of Spacelink. I feel that the lessons and activities when used will help increase the number of scientists in the nation, and will make our entering college students better prepared for college work.

NASA, an agency dedicated to achieving excellence, has for many years contributed to excellence in education in the nation's schools. The material prepared for Spacelink is for use in Grades K-12 which is in keeping with NASA's effort and is available to anyone within and outside the United States. NASA is hereby applauded and encouraged to continue this worthwhile endeavor.

REFERENCES

1. Career Opportunities in Aerospace Technology, NN-100
NASA, Marshall Space Flight Center, Huntsville, Alabama
2. NASA Aerospace Education Services Project
Oklahoma State University, Stillwater, Oklahoma
3. NASA Facts, NF-150/1-86, Marshall Space Flight Center,
Huntsville, Alabama
4. NASA Spacelink, Marshall Space Flight Center,
Huntsville, Alabama
5. NASA Educational Workshop for Elementary School
Teachers (NEWEST), Marshall Space Flight Center,
Huntsville, Alabama
6. Tennessee Space Week TEA * NASA Lesson Plan
Tennessee Lesson Plans #20, Rocket Launch System
Activities

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

SIRE - a Simple Interactive Rule Editor for *NICBES* .

Prepared by:	Alex Bykat
Academic Rank:	Professor
Institution:	University of Tennessee at Chattanooga
Department:	Center of Excellence for Computer Applications
NASA/MSFC:	
Laboratory:	Information and Electronic Systems
Division:	Electrical
Branch:	Electric Power Branch
MSFC Colleague:	David J. Weeks
Date:	August 8, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

SIRE - a Simple Interactive Rule Editor for NICBES .

Alex Bykat

Center of Excellence for Computer Applications

University of Tennessee,

Chattanooga, TN 37403

ABSTRACT

To support evolution of domain expertise, and its representation in an expert system's knowledge base, a user-friendly rule base editor is mandatory. NICBES, a prototype of an expert system for the Hubble Space Telescope power storage management system, does not provide such an editor. In the following, we present a description of SIRE - a Simple Interactive Rule Base Editor for NICBES.

SIRE provides a consistent internal representation of the NICBES knowledge base. It supports knowledge presentation and provides a user-friendly and code language independent medium for rule addition and modification. SIRE is integrated with NICBES via an interface module. This module provides translation of the internal representation to Prolog-type rules (Horn clauses), latter rule assertion, and a simple mechanism for rule selection for it's Prolog 'inference engine'.

Acknowledgements.

I would like to express my appreciation of the hospitality extended during my last two summers spent at Marshall Space Flight Center. I have enjoyed and benefitted from participation in the Summer Fellowship Program. Much of this is due to the efforts of the program coordinators, Mrs. E. Cothran (MSFC) and Dr. M. Freeman (The University of Alabama).

Working in the Electric Power Branch was an enjoyable experience. This was due in particular, to Mr. David J. Weeks' hospitality, warm reception and friendly guidance through the maze of MSFC offices. Thank you Dave! My thanks go also to Dr. G.R. Wallace for both, putting me in touch with Dave and for supporting our cooperation, and to Mr. J.L. Miller, Mr. W.G. Shields and Mr. J.R. Lanier for supporting my stay in their division.

Introduction.

A prototype of NICBES -- Nickel Cadmium Battery Expert Systems -- is currently operational and under testing on the HST test bed. The prototype is written as two separate subsystems: the data handler, and the diagnosis system.

The data handler subsystem is written in Microsoft C language. Its main function is to receive telemetry data from the test bed, and to 'massage' this data into a form suitable for input to the diagnosis system.

The diagnosis subsystem is written in Arity Prolog language. Its main function is to take the data prepared by the data handler, and evaluate this data to discover any exception situations that might be indicated. In particular, the diagnosis tries to evaluate the state of the batteries (as indicated by the data) for possible malfunctions (charge leakages, overheating, etc), and for maintenance operations (recharging, etc).

The diagnosis subsystem displays its findings in terms of graphs indicating various 'trends' of batteries performance, as well as alarm messages (when they are necessary). The graphical operations needed for the trend displays were written in Microsoft C.

The NICBES prototype performs well, but a number of problems have been exhibited during its test runs. In addition, the experience with the prototype has provided insights which allowed to identify some needed additions and improvements of the system. Some of these problems and needed improvements are listed below.

NICBES problems:

1. Battery plots show incorrect data points. This apparently occurs
 - a. consistently in the last data point, and
 - b. intermittently in other positions.

The first case is probably due to NICBES incorrectly handling the last data point (interrupted orbit?).

The second case may be due to bad data points. If this is the case the origin of the bad data has to be traced. A number of possibilities exist:

- A. NICBES data handler garbles the data.
- B. NICBES data handler gets confused because of garbled header in the data package.

The data handler receives transmission once every minute. Each transmission lasts precisely six seconds. An orbit telemetry data consists of 96 such transmission bursts, with each burst consisting of a data header followed by 370 telemetry values. A data header consists of a special start-data character -- the character A -- and is followed by nine data items, which identify the history of each data burst.

To receive the transmission, the data handler looks for this special start-data character. If the character A is not recognized, the data is rejected as noise. When the start data character is recognized, the data handler starts collecting the data and stores it as orbit data files, for a subsequent use by NICBES.

Herein lies a possible source of trouble -- here is a situation where two wrongs can make a right! A misread start-data-character will cause valid data (and in particular the header) to be rejected. A subsequent data item, misread as a start-data-character, may cause acceptance of the remaining part of a transmission as a complete telemetry. Of course, this latter data will be out of phase, and may very well cause bad data points to appear on plots.

To verify this scenario, a software filter should be written. This filter will be applied to (suspect) archived telemetry data files to detect short transmissions. Once the relationship between these faulty files and the bad data points is verified, the filter can be integrated with the data handler, to reject such faulty transmissions.

- C. Telemetry data contains bad values. This case should raise an alarm prior to transmission of data to NICBES. Since no alarm is raised there are two 'sub-possibilities':
 - C1. alarm does not work (eg. it does not recognize invalid data), or
 - C2. alarm is not recognized by NICBES
- 2. NICBES evaluates batteries performance from data collected over 12 orbits only.

Battery performance trends should be evaluated over a large number of orbits. However, increase in the number of orbits is currently infeasible due to the hardware and software environment of NICBES prototype.

NICBES improvements:

- 1. Presentation of trends over a long period of time. This requires ability to handle more than the 12 orbits in the current version. The number of orbits to be aimed for is 400 to 500.

Before we can look at a technical solution to this problem, we have to change NICBES environment. The 12 orbits limitation is due to the NICBES hardware environment which consists of an IBM AT with 640 KB memory and 20 MB hard disk running under DOS.

- 2. Knowledge Base Rule Editor. It is desired to have a user friendly facility for display and modification of NICBES rules. Currently, to change or even to display a rule, Prolog programming knowledge is required.

This should be doable over the Summer, provided the NICBES has a well defined knowledge representation. If this latter holds, at least a rudimentary Rule Editor can be build incorporating functions such as

- a) display a rule,
 - b) delete a rule
 - c) add a rule
- 3. Graphic representation of battery voltage uses a 'flexible spread' range and computed scale within that range. This makes visual recognition and comparison rather difficult for a human eye.

The scale have the same range, those the origin point of the scale can be adjusted to represent the (minimal) data value.

- 4. Multitasking is necessary to continue data collection when NICBES consultation is in progress.

Desqview was suggested as a quick fix to add the multitasking capability. However, it is my opinion (based on my current and superficial knowledge of DV) that this software will not provide multitasking within one package. I have to investigate that further, but I believe that DV provides 'multiple window multitasking' ie. time sharing of applications, not multitasking within application.

- 5. Printout of messages without interrupting NICBES data collection.

This is related to the above requirement.

The above list of needed improvements is arranged in order of priority. However, not all of these improvements can be tackled with currently available hardware configuration. In particular, presentation of trends over a long period of time (improvement 1) requires a large main storage capacity, certainly larger than the 640 Kbyte available on the currently available NICBES computer (IBM PC/AT). Similarly, the multitasking capability (improvement 4) is not available for the IBM PC/AT. Thus, these improvements will be

relegated to implementation of NICBES on an I80386 class of computer.

The requirement for a knowledge-base rule editor can be developed for the current NICBES prototype. It is this new subsystem that is described in the subsequent pages of this report.

Objectives.

The main objective of the following work is to provide the NICBES prototype with a subsystem which supports acquisition and modification of nickel-cadmium battery management rules. This subsystem will be used by the (human) battery management experts to extend and to formalize their experience (with the said batteries) into a knowledge base. This expert knowledge base is captured in the form of production rules, and is used by the (software) expert system -- NICBES -- in diagnosis of battery performance.

To achieve this objective, it was necessary to write a software system -- SIRE -- which provides:

1. Internal representation of knowledge in a consistent manner.
2. Capability to display the captured domain knowledge.
3. Capability to add new domain knowledge.
4. Capability to modify captured knowledge in a programming code independent manner.
5. Interface to NICBES to integrate the subsystem and to avail the knowledge base for subsequent diagnosis of battery performance.

SIRE - Simple Interactive Rule Editor.

NICBES was intended to prove the capability of expert system technology in (eventually) autonomous management of the power supply system for the Hubble Space Telescope. Consequently, and as this technology dictates, NICBES has two main components:

- 1) domain knowledge, and
- 2) mechanism to manipulate this knowledge.

Building the domain knowledge demands consideration of two aspects:

- 1) knowledge acquisition and
- 2) knowledge representation.

Of course, the latter is influenced by the fashion in which the knowledge manipulation mechanism will use it.

Evolution and changes in domain expertise, mandate from an expert system provision for a programming-language independent rule editing. Current version of NICBES, does not support such capability. The rules of NICBES are simply programmed in Prolog, demanding familiarity with Prolog to effect rule changes. Since this essentially requires a new and highly technical skill from the domain experts, it is not a satisfactory situation. Indeed, to change a rule, in addition to knowing Prolog, the domain expert would have to be familiar with NICBES internal code structure. To alleviate this situation we shall implement a 'friendly' rule editor -- SIRE. Figure 1 shows the SIRE environment within NICBES.

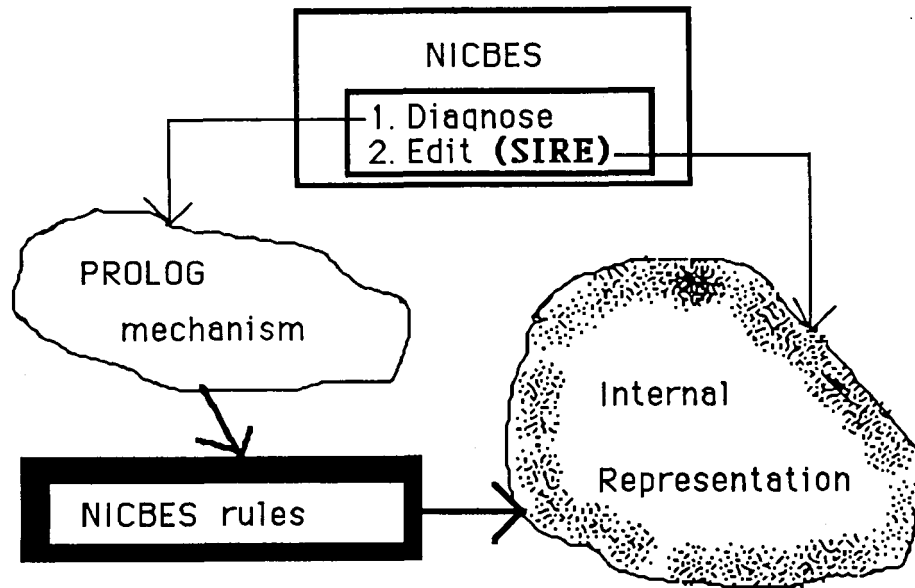


Fig.1. SIRE environment

Rule representation.

As stated previously, current version of NICBES does not represent rules in a consistent and well defined way. The rules are simply programmed in Prolog, with structure depending on each individual rule type. The following is an example of a NICBES programmed rule:

```
advice(Bat,1):-
    write($** ADVICE ON RECONDITIONING BATTERY: **$), NL,
    get_one(Trend1,Trend2,Trend3,Trend4,Trend5,Bat),
    recond(Trend1,Trend2,Trend3,Trend4,Trend5).
```

```
get_one(Trend1,Trend2,Trend3,Trend4,Trend5,Bat):-
    .....
```

```
recond(T1,T2,T3,T4,T5):-
    write($EOD voltage trend is$), write(T1), nl,
    write($In-charge voltage trend is$), write(T2), nl,
    write($Recharge ratio trend is$), write(T3), nl,
    write($EOD divergence trend is$), write(T4), nl,
```

```
write($In-charge divergence trend is$), write(T5), nl,  
tell_1(T1,T2,T3,T4,T5).
```

```
tell_1(strongly_down,_,_,strongly_up):-  
    write($Reconditioning advised to correct failing  
    capacity.$),nl.
```

In the above, the *advice* predicate invokes the diagnostic evaluation of battery *Bat* performance data. The second argument in *advice* indicates a choice of the type of diagnostics to be performed. (Currently, there are three choices possible: reconditioning, charging regime, and workload.)

The *get_one* predicate uses orbit telemetry data files to calculate the battery *Bat* trends. These trend values are then "fuzzed" into various intervals, and the latter are given symbolic names such as: *strongly_down*, *down*, *up*, *strongly_up*. Note that *get_one* calculates variables which are then used as initial parameters for the diagnostic rules.

The *recond* predicate takes these symbolic trend names, and displays their values as an overall battery assessment. Thus, *recond* merely reports the trend values, and then passes control to *tell_1*.

Finally, the *tell_1* predicate uses the appropriate symbolic trend values to display NICBES recommendations in English. (Although not shown above, there are altogether 11 different recommendations that can be produced by *tell_1*. of course, different recommendations are produced in response to different symbolic trend values.)

To support SIRE functions, the rules (such as above) have to be structured and represented consistently. The following describes the knowledge representation that shall be adopted for SIRE (and hence NICBES) rule representation.

The internal representation of NICBES rules will consist of the five parts:

- 1) rule category
- 2) rule number
- 3) parameter initialization

- 4) premise specification
- 5) conclusion specification.

These five parts will be presented as five arguments of the PROLOG predicate rule/5, eg:

```
rule(Category, Number,  
      params(Par_List), premis(Pre_List), conclude(Con_List)).
```

Note that the rule category allows us to structure the rule base into rule groups corresponding to each category. This allows increase in search efficiency during the inference process. Within each category, a rule is identified via its number. Globally, each rule will be uniquely identified by the tuple (rule category, rule number).

1) **Rule category** names are created by the user. Current NICBES rule base will be divided into categories such as reconditioning, charge and load. Other rule categories can be established by the SIRE user.

2) **Rule number** will be used to identify a rule within a rule category. Rule numbers must be unique within a category, though they may be repeated across the categories.

3) **Parameter initialization** list `Par_List` will consist of elements, each in the form of a valid Prolog term. Typically, these terms will represent assignment of a value to a parameter which is used subsequently in the rule body. The assignments are of the form:

Parameter:= Value

where

Parameter -

specifies the variable to be used in the rule. This is the actual slot that will contain the the parameter's value.

Value -

specifies the initialization procedure for the parameter. The value can be of three types:

ask-

specifies that a value for this parameter should be asked for interactively from the SIRE user.

get_trend (Type,Bat),
get_average (Type,Bat)-

specifies that the function be evaluated and its return value be assigned to the parameter.

<other> -

specifies that *<other>* be assigned as entered to the parameter's value.

4) **Premise specification** list *Pre_List* will consist of elements, each of which is a Prolog term. The terms specified in the list, combined with the AND operator (Prolog's comma) will constitute the antecedent of the rule.

5) **Conclusion specification** list *Con_List* will consist of elements of the same form as those in the premise specification list. The terms in *Con_List* will be executed subject to satisfaction of the rule's premises.

Rule interpretation.

The following is an example of a NICBES rule, stated in its (new) internal representation. This internal representation is followed by it's English interpretation.

```
rule( recondition,
      1,
      params([
        $uses(Bat)$,
        $EOD_VOLTAGE_TREND:=get_trend(eod_voltage,Bat)$,
        $IN_CHARGE_DIVERGENCE_TREND:=
        get_trend(in_charge_divergence,Bat)$]),
      premis( [
        $EOD_VOLTAGE_TREND = down$,
        $IN_CHARGE_DIVERGENCE = strongly_up$]),
      conclude([
        $write('reconditioning is advised')])).
```

English translation:

Parameter initialization:

Get the global value *Bat*

Evaluate *EOD_VOLTAGE_TREND*.

Evaluate *IN_CHARGE_DIVERGENCE_TREND*.

The rule:

If

EOD_VOLTAGE_TREND is down and

IN_CHARGE_DIVERGENCE_TREND is strongly_up

Then

write: *reconditioning is advised*.

Note that the elements of the three rule parts are specified as Prolog strings. This is necessary due to Arity Prolog's failure to preserve symbolic names of variables. In Arity Prolog implementation, variable's symbolic name is replaced with a coded name. Although this saves some internal storage, information conveyed by the symbolic name (so vital to program's documentation) is lost at run time. (This is particularly bothersome in debugging stage.)

SIRE capabilities.

SIRE capabilities are accessed via two level menu system. The selections from the two levels are combined providing an (almost) Cartesian product of choices, as indicated in the table below:

First Level Menu		show	delete	add	change	statistics	print	quit
Second Level Menu	<category>	×	×	×	×	×	×	×
	rule	×	×	×	×	×	×	×
	class	×	×			×	×	×
	all	×	×			×	×	×
	quit	×	×	×	×	×	×	

SIRE functions

Note that <category> will be actually presented as a list of category names, eg: recondition, charge, load, etc.

SIRE design.

SIRE is written in Arity Prolog Version 4. The editor consists of

1. menu module,
2. rule base module,
3. editor functions module,
4. rule interface module,
5. on-line help module, and
6. SIRE control module.

Menu module.

The menu module consists of SIRE menu definitions. These are statements using the Arity Prolog Screen Toolbox. The menu definitions are kept in file MENUS.ARI which is consulted by the SIRE control module.

As stated, SIRE capabilities are accessed via two level menu system. The first level affords selection of requests for NICBES rule-manipulation functions such as:

show, delete, add, change, save, statistics, help,
print, quit.

Selection of any option (other than the quit option) will result in display of a second level menu. The second menu level allows selection of NICBES rules to be manipulated. The choices provided in the second menu level are:

<category>, rule, class, all, quit.

Here, <category> represents a list of names of rule categories that actually exist currently in the rule base. The <category> list of names in the menu is updated dynamically, so that it can be looked at as a dynamic test for rule class changes.

Rule base module.

The rule base module contains the internal representation of NICBES diagnostic rules. The syntax of the internal representation is quite simple:

```
RULE      ::= ule(CATEGORY, NUMBER, PARAMETER, PREMISE,  
                  CONCLUSION)  
CATEGORY  ::= <user defined name>  
NUMBER    ::= <integer>  
PARAMETER ::= params( TERM )  
PREMISE   ::= premis( TERM )  
CONCLUSION ::= conclude( TERM )
```

where TERM is a string written in syntax as defined by the Arity Prolog term definition. Generally speaking, this means a syntax defined by

```
TERM      ::= <operator> ( ARG_LIST )  
TERM      ::= <atom> | <number> | <string> | <variable>  
ARG_LIST  ::= TERM  
ARG_LIST  ::= ARG_LIST, TERM
```

Note that, as is usual in Prolog, the operators can be written in prefix or infix notation as appropriate.

Rule interface module.

The interface module is provided to integrate the new internal rule representation with the NICBES diagnostic system. The main function of this module is to translate the internal rule representation into a Prolog clause form. This 'executable' form is then supplied to NICBES inference mechanism for evaluation.

The interface module performs as a three stage module. In the first stage the internal form of SIRE rule is retrieved and then translated into the Prolog expression form. It is at this stage that the syntax of the rule is checked -- a rule written in valid syntax is one that conforms to the syntax of a Prolog expression. If the rule is found to be syntactically invalid, a suitable message will be output, and the translation of the rule is abandoned. The user can then use SIRE to correct the rule.

In the second stage a name for the rule, based on the rule's CLASS and NUMBER is generated. The name is then prefixed to the rules body, and the valid clause is then asserted (added) as a rule. Thus a rule now appears as a valid Prolog clause. While asserting such a rule, any existing rule with the same name is retracted (erased). Translation and assertion of the rules into NICBES rule base is done only once -- the first time a consultation for the required class is requested. This minimizes the internal representation overheads due to the translation process.

In the third stage the Prolog inference mechanism is invoked with the asserted rules supplied as the rule base. This process consists, essentially, of evaluation of rules in the order of their assertion

sequence. If a rule succeeds, the result is exhibited to the user; if the rule fails, another rule is extracted from the rule base and tested.

While a rule is evaluated, this module retrieves the global values indicated in the parameter initialization part and makes them available to the rule. In addition, the 'basic NICBES operations' are evaluated (as indicated) to provide the initial values to the rule variables. (These basic operations, eg. *get_trend*, *ask*, *:=*, etc., are defined in this module.)

Since many rules may require evaluation of the same operation for the same battery using the same data, there is a considerable opportunity for inefficiency. To avoid this, each operation which obtains new information, generates a 'lemma' which is then used by the repeated calls without a need to repeat the 'proof'. Thus, for example, a new information obtained by *get_trend(charge,1)* will be asserted as a new fact; next time the *get_trend(charge,1)* is called, this fact is then simply retrieved rather than recomputed.

Editor functions module.

This function provides the capabilities for modification of the internal knowledge representation. The editor functions currently provided include:

show, delete, add, change, save, statistics, help,
print, quit.

On-line help module.

SIRE has an extensive on-line help available to its user. The on-line help describe all of the major functions of SIRE, and is accessible via the *help* selection in the first level menu. The capability of the on-line help is further indicated in the following figure.

First Level Menu						
	show	delete	add	change	statistics	help
help	X	X	X	X	X	X

SIRE online help

Control module.

SIRE control module provides access to the functions of the editor. This module calls the menu modules to display the menus, and subsequently combines the user responses to determine which of the editor functions are to be activated by the editor function module. On completion of user requests, the control module performs maintenance operations on behalf of the rule module.

Conclusions.

The objectives of this project have been successfully met. The SIRE subsystem has been successfully integrated with the NICBES prototype, and will provide it's users with a system for display of the knowledge base status, addition of new knowledge in the format of SIRE/NICBES rules, and modification of these rules as the dynamically evolving expertise may require.

In addition to the above, a rudimentary diagnosis explanation facility has been implemented. It provides not only display of conclusions, but also identification of rules which were instrumental to reaching the conclusion. Although this explanation facility should be further developed, it will already be of benefit by providing a means to pinpoint the rules and to focus on the knowledge which may need revision.

All of these functions are now 'user-friendly'. Thus, the user of NICBES no longer requires knowledge of programming in Prolog, nor does he require knowledge of the internal structure of NICBES code.

As a by-product of this work, it has been ascertained that the NICBES prototype code must be rewritten in order to function as a useful product. This conclusion is inevitable when the sluggishness of NICBES is compared with its current 'size parameters'. Thus, although NICBES has only some 40 rules in its knowledge base, and refers to only 12 orbits, it is already heavily i/o bound. The main reasons for the frantic disk activities exhibited by NICBES, are due to the DOS limitation of 640 KB for memory, the size of the Prolog system, and the storage cost of recursive calls in Arity Prolog.

The slow-down due to the excessive disk activities is further compounded by the battery performance trends and averages calculations. For example, although only 12 orbits are taken into account in calculation of a performance trend, such calculation takes about 20 seconds. With five trends to be calculated, the user has to wait for almost two minutes. Now for the crunch: if number of orbits is increased to 1000, the calculations (as programmed now) of five trends will take 3 hours!!

References.

Martin Marietta Corp. "Final report for NICBES", MCR-85-641, 1986

Martin Marietta Corp. "User manual for NICBES", MCR-86-673, 1986

Martin Marietta Corp. "Requirements specification for NICBES", MCR-86-674, 1986

Martin Marietta Corp. "Program maintenance manual for NICBES", MCR-86-675, 1986

The Arity/Prolog Programming Language, (Version 4), 1986

The Design Arity Screen Toolkit, 1986

Bykat, A. "User manual for SIRE", 1988

Bykat, A. "Expansion of NICBES capabilities", MSFC, NAG8-105, 1988

N89-21731

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

**MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA**

SMALL EXPENDABLE DEPLOYER SYSTEM MEASUREMENT ANALYSIS

Prepared by:	Connie K. Carrington
Academic Rank:	Assistant Professor
University and Department:	University of South Carolina Mechanical Engineering Department
NASA/MSFC: Laboratory: Division:	Program Development Orbital Systems
MSFC Colleague:	Charles C. Rupp
Date:	August 5, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

SMALL EXPENDABLE DEPLOYER SYSTEM MEASUREMENT ANALYSIS

by

Connie K. Carrington
Assistant Professor of Mechanical Engineering
University of South Carolina
Columbia, South Carolina

ABSTRACT

The first on-orbit experiment of the Small Expendable Deployer System (SEDS) for tethered satellites will collect telemetry data for tether length, rate of deployment, and tether tension. The post-flight analysis will use this data to reconstruct the deployment history and determine dynamic characteristics such as tether shape and payload position. Linearized observability analysis has determined that these measurements are adequate to define states for a two-mass tether model, and two state estimators have been written.

ACKNOWLEDGEMENTS

The summer program participant would like to thank Program Development at Marshall Space Flight Center for a most enjoyable summer of research. The SEDS analysis provided a "real-life" problem that has been educational to the investigator, and will be discussed in her estimation course. In particular, I would like to thank Chris Rupp and Jim Harrison for all their help in defining the problem and their advice in addressing it.

The investigator would also like to acknowledge the excellent summer program that Mike Freeman and his staff have created. It was a pleasure to be a part of this program.

INTRODUCTION

The first flight of the tethered Small Expendable Deployer System (SEDS) will be launched from a Delta II. A 14 kg payload will be deployed downward to a tether length of 2 km, where it will librate to vertical and the tether will be cut.

The data system will record the number of reel-turns as the tether is deployed, and tether tension. These measurements will be downlinked to a ground station so that deployment can be monitored, and the data will be stored for post-flight analysis.

The telemetry data will provide measurements of tether length, rate of deployment and tether tension. To determine if these measurements are adequate for reconstruction of the deployment dynamics, system observability calculations have been made (based on a constantly-updated linear model), and two state estimators have been developed.

OBJECTIVES

The objectives of the summer faculty research were to:

1. Determine if the turns-count and tension measurements are adequate to reconstruct the deployment dynamics.
2. Determine if the measurement sampling rate is adequate.

STATE MODEL

Several computer simulations of tether deployment dynamics are available, ranging from planar simple-pendulum representations to three-dimensional partial-differential-equation models. The summer investigator chose Energy Science Laboratories (ESL) BEADSIM model to provide the tether dynamics state equations, since it is relatively simple and yet still produces results that are comparable to more complex models. BEADSIM is a lumped mass model, in which masses or "beads" are added as the tether becomes longer. No out-of-orbit-plane motion is modelled, and the external forces on each bead are the gravity gradient, aerodynamic drag, and Coriolis and centripetal accelerations. The equations are written using a Cartesian coordinate frame with an origin at the center of mass and moving at orbit speed (Fig. 1).

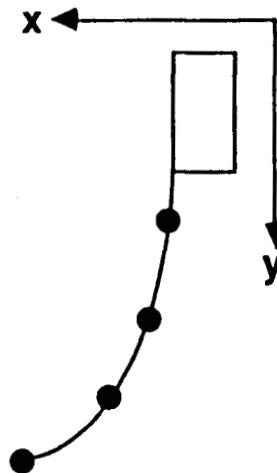


FIG. 1. BEADSIM Tether Model.

Each bead's motion is governed by a second-order differential equation, with a uniform tether tension providing the coupling between beads. Four states represent the motion of each bead: the x and y positions, and the x and y velocities. Deployment

characteristics such as tether shape, vibration amplitudes, and payload position and velocity are determined as the states change in time.

Simulations like BEADSIM are dependent on initial conditions for the states; any changes in the initial conditions will produce a different deployment trajectory (Fig. 2). The SEDS measurements would not be necessary for post-flight analysis if the initial positions and velocities of the deployer and payload were known exactly, and if parameters such as the aerodynamic drag coefficient were accurate. However, variations in the initial conditions greatly effect tether deployment time, for example, so that an estimate of initial conditions is not adequate to determine deployment characteristics.

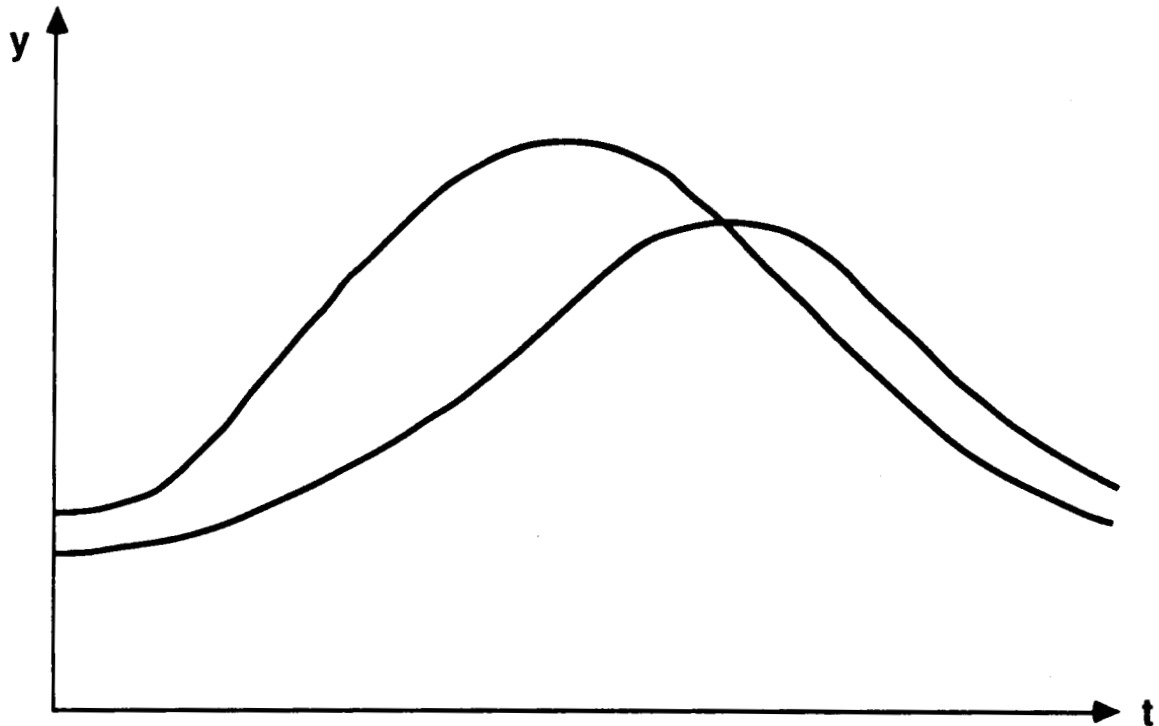


FIG. 2. Changes in the Initial Conditions Produce Different Deployment Trajectories.

The SEDS measurements consist of tether length l , length rate \dot{l} , tension T , and the Delta second stage

position and velocity. These measurements are nonlinear functions of the states, represented by the following measurement equations

$$l = G(y) \quad (1)$$

where l represents the vector of measurements at a given time, and y the corresponding states. The BEADSIM state equations are also nonlinear, and may be written as follows

$$\dot{y} = F(y) \quad (2)$$

The measurement and state equations will be used to determine system observability and to develop the state estimators.

SYSTEM OBSERVABILITY ANALYSIS

Given state and measurement equations like those in (1) and (2), a system is totally observable if all states can be determined from the measurements. For linear systems, a full-rank observability matrix ensures system observability. Since our system is nonlinear, a linear approximation based on the first term of a Taylor series will be used, so that the measurement equations (1) become

$$l = Hy \quad (3)$$

and the state equations become

$$\dot{y} = Ay \quad (4)$$

where H and A are the following matrices

$$H = \frac{\partial G}{\partial y} \bigg|_{y_{\text{current}}} \quad (5)$$

$$A = \frac{\partial F}{\partial y} \bigg|_{y_{\text{current}}} \quad (6)$$

Closed-form expressions for the elements in these matrices were derived from the functions F and G , so that errors from numerical differentiation were avoided. Since these matrices are evaluated at the current values for the states y , they will change as the states change. Hence system observability based on these matrices must be checked at each time step.

The observability matrix is defined as

$$O = \begin{bmatrix} H \\ HA \\ \vdots \\ HA^{n-1} \end{bmatrix} \quad (7)$$

where n is the number of states to be estimated. A is an 8×8 matrix for eight states corresponding to two beads, and H is a 7×8 matrix for the seven measurements l , \dot{l} , tension T , and the x , y position and velocity of the Delta. The observability matrix O is 56×8 . The rank is the number of independent rows in O , and is calculated using the singular value decomposition. The number of nonzero singular values at each time step was always eight, indicating that the linearized system for two beads is totally observable. Hence the positions and velocities of the tether deployer and payload may be determined by these measurements.

The rank of the observability matrix was also calculated without the tension measurement, and the system was still totally observable. This analysis indicates that tension measurements are not necessary to reconstruct the tether dynamics using two beads.

STATE ESTIMATORS

The measurements are not the same variables as the states, so that deployment characteristics such as tether shape cannot be directly determined from the measurements. Furthermore, the measurement process adds noise. A state estimator will use the measurements as input, filter the measurement noise, and then output the values of the states. Two estimators have been written: a batch estimator that finds the best initial conditions to minimize measurement errors, and a Kalman estimator that uses the current measurements to estimate the states. Both estimators process simulated measurements generated by BEADSIM to which white noise was added. The standard deviations of the noise are listed in Table 1.

TABLE 1 Standard Deviations for Measurement Noise

Measurement		σ
length	l	0.01 m
length rate	\dot{l}	0.001 m/s
tension	T	0.01 N
Delta	$\begin{cases} x_d \\ y_d \end{cases}$	1.0 m
position	$\begin{cases} \dot{x}_d \\ \dot{y}_d \end{cases}$	1.0 m
Delta	$\begin{cases} \ddot{x}_d \\ \ddot{y}_d \end{cases}$	0.001 m/s
velocity	$\begin{cases} \ddot{x}_d \\ \ddot{y}_d \end{cases}$	0.001 m/s

Least-Squares Batch Estimator

The least-squares batch estimator processes a batch of measurements to estimate initial conditions. The user inputs approximate initial conditions y_0 for the states, and the estimator integrates forward over the number of time steps for which measurements are available. The integrated values for the states at each time step are used to calculate measurement values \hat{l} for tether length, length rate, etc., which are then compared to the actual measurements l . The error vector $\{l - \hat{l}\}$ between the actual and calculated measurements is multiplied by a gain matrix that minimizes the errors in a least-squares sense, and the result is added to the

guess for the initial conditions y_0 . This procedure is based on Gaussian differential correction, and is repeated until the measurement errors $\{1-\hat{1}\}$ are small enough. The current iterate for the initial conditions is assumed to be correct, and the last set of integrated state values represent the true states.

Since BEADSIM adds more states as the tether is deployed, the measurements can be divided into batches that change when a bead is added. When enough tether has been deployed to add a third bead, the current batch of measurements is processed to determine initial conditions for the first two masses. The next batch of measurements is used to determine initial conditions for the third bead, and the process continues until all beads have been added.

Estimators for dynamic systems use the linearized system's transition matrix to define the relationship between initial conditions and later state values. The state transition matrix $\Phi(t, t_0)$ is the transformation that takes the initial state y_0 into the later state y

$$y(t) = \Phi(t, t_0)y(t_0) \quad (8)$$

For linear systems, the state transition matrix obeys the same differential equation as the states themselves

$$\dot{\Phi}(t) = A\Phi \quad (9)$$

and can be integrated forward in time along with the state equations. For nonlinear systems, however, equation (9) is a linearized approximation valid only in a neighborhood of the current state value. As the integration progresses and the current state gets further from the initial conditions, the approximation becomes inaccurate and the state transition matrix does not produce the same state values as the integrated state equations. This inaccuracy reduces the ability of the least-squares batch estimator to process measurements far from the initial conditions. Although the batch estimator appeared to estimate the correct initial conditions for two-bead tether simulations, the

principal investigator feels that a Kalman state estimator is more accurate.

Extended Kalman Estimator

Kalman filters are based on linear system theory, but can be used with an integrator to extend the algorithm to nonlinear processes. Unlike the batch estimator, measurements are processed one-at-a-time, so the current set of measurements are used to estimate the current state.

As an example, consider a Kalman estimator that processes a measurement l and estimates a state y . An integrator predicts an estimate of the state y_1 at time t_1 , and that estimate is used to calculate a value for the measurement \hat{l} . The actual measurement l is compared to \hat{l} , and the error $\{l - \hat{l}\}$ is multiplied by a gain G . $G(l - \hat{l})$ is the correction that is added to y_1 to give a better estimate of the state at time t_1 . The integrator goes forward another time step using the corrected state value at time t_1 , and a new correction is calculated from the next measurement.

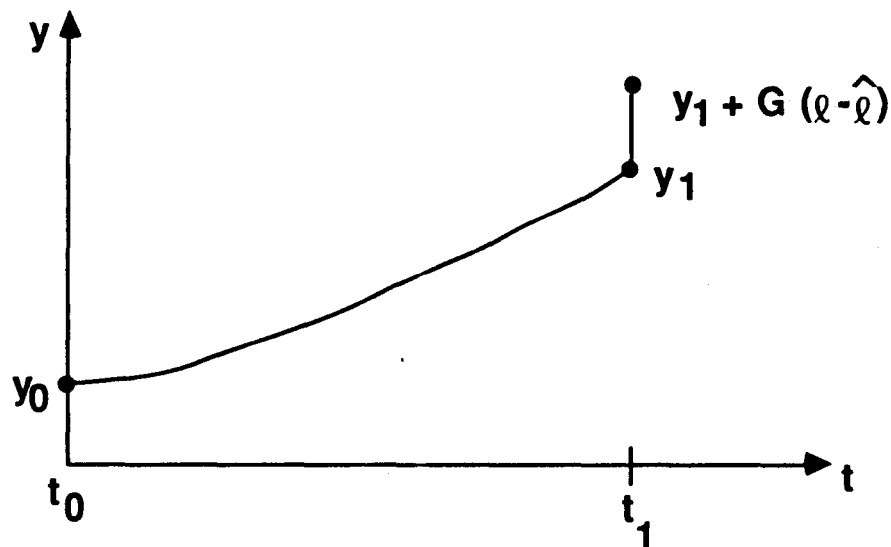


Fig. 3. Kalman Estimator Process.

The gains G are calculated by a minimum variance estimation algorithm based on assumed measurement noise statistics. Process noise statistics may also be included, so that the measurement noise and process noise become adjustments to tune the filter.

The extended Kalman filter uses the full nonlinear state equations to predict the next state, but still relies on the linearized state model for the state transition matrix, and a linearized measurement equation. Unlike the batch estimator, the Kalman filter only needs the state transition matrix that takes the current state to the next state

$$y(t_2) = \Phi(t_2, t_1)y(t_1) \quad (10)$$

so that the linearizing approximations that made the batch estimator inaccurate do not occur. As a check, the estimator can multiply the current state $y(t_1)$ by the state transition matrix and compare the result with that produced by integration of the nonlinear equations. Throughout the deployment, the two-bead state values agree to at least four decimal digits using a one-second time step, indicating that this time step is small enough to accommodate any linear approximations in the Kalman filter.

Several algorithms are available for coding the Kalman filter. The traditional filter algorithm produced overflows, so the U-D square-root factorization algorithm was programmed. This coding is more stable according to the published literature, but although it did not overflow it could not accurately estimate the tether angle during deployment. It was also extremely sensitive to assumed noise levels that tune the filter. Since the U-D algorithm could not produce satisfactory results, the traditional algorithm was modified to include process noise, and with this modification it successfully estimated the (two-bead) tether dynamics throughout deployment. The maximum error between the "true" states from BEADSIM and the estimated states was 10%, and occurred in the vertical position of the endmass. This error results from using only two-beads

in the estimator, as compared with the 21 beads used the generate the "true" states. A slight reduction in error was achieved by adjusting the masses of the two beads as the tether deployed, but a multibead capability must be added to the estimator to further reduce errors.

CONCLUSIONS AND RECOMMENDATIONS

The system observability analysis has indicated that tether length, length rate, and Delta position and velocity measurements are adequate to estimate positions and velocities of a two-bead tether model. The Kalman filter produces estimates within 5% of the "true" states for the first three-quarters of deployment, with a maximum of 10% error when fully deployed. Both the observability analysis and the estimators indicate that tension measurements are not necessary; noise levels and error bias associated with tension measurements may actually degrade the estimation process. The data sampling rates proposed at this time appear to be adequate to produce measurements at the one-second time interval needed by the Kalman estimator.

Recommendations for improvement include the addition of software to add more beads in the estimation process. The partial derivatives for the A and H matrices of equations (5) and (6) must be derived for multiple beads, and mechanisms developed for adding columns and rows to the gain and covariance matrices when beads are added. This improvement will reduce errors as well as add the capability to estimate flexible tether shape and vibrations.

The measurement noise added for simulation purposes was not as "white" as desired, due to correlation in the Turbo-Pascal random number generator. Noise models that are statistically more correct should be used, and accurate noise models for the tether length and length rate measurements should be developed. This would entail the development of a tether-reel dynamics model that would translate the turns-count data into length and length-rate measurements. The model should reflect the periods and biases associated with changes in the winding directions.

An alternate system model using polar coordinates l and θ rather than the Cartesian coordinates of BEADSIM may produce a more accurate estimator. The tether length and length rate measurements will then be linear functions of the states, which may reduce current errors due to linearization.

REFERENCES

Bierman, Gerald J., "Measurement Updating Using the U-D Factorization," Proceedings of the 1975 IEEE Conference on Decision and Control, pp. 337-346.

Bierman, Gerald J. and Thornton, Catherine L., "Numerical Comparison of Kalman Filter Algorithms: Orbit Determination Case Study," Automatica, Vol. 13, 1977, pp. 23-35.

Carroll, Joseph A., "Tether Applications in Space Transportation," Acta Astronautica, Vol. 4, 1986.

Carroll, Joseph A., "SEDS Data System Functional Requirements," report to MSFC Program Development, May 1988.

Carroll, Joseph A., "The Small Expendable Deployment System (SEDS)," presented at the PSN/NASA/ESA Second International Conference on Tethers in Space, Venice, Italy, October 1987.

Carroll, Joseph A. and Alexander, Charles M., "SEDS, The Small Expendable-tether Deployment System," Final Report NASA SBIR Phase II, NAS8-35256, December 1987.

Control Dynamics Company, "Evaluation of Tether Dynamics and Control System Interaction," NASA Final Report, NAS8-34667, June 1983.

Junkins, John L., An Introduction to Optimal Estimation of Dynamical Systems, Sijthoff & Noordhoff International Publishers, Alphen aan den Rijn, The Netherlands, 1978.

Kohler, P., Maag, W., and Wehrli, R., "Dynamics of a System of Two Satellites Connected by a Deployable and Extensible Tether of Finite Mass," NASA Final Report, October 1978.

Press, William H., Flannery, Brian P., Teukolsky, Saul A., and Vetterling, William T., Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, 1986.

Wolovich, W.A., Linear Multivariable Systems, Springer-Verlag, New York, 1974.

N89-21732

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

**MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA**

**The Role of Radiation-Dynamics Interaction
in Regional Numerical Weather Prediction**

Prepared by:	Chia-Bo Chang
Academic Rank:	Assistant Professor
University and Department:	Texas Tech University Atmospheric Science Group
NASA/MSFC:	
Laboratory:	Structures and Dynamics
Division:	Earth Science and Applications
Branch:	Fluid Dynamics
MSFC Colleague:	Franklin R. Robertson
Date:	August 8, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

The Role of Radiation-Dynamics Interaction in Regional Numerical Weather Prediction

by

**Chia-Bo Chang
Atmospheric Science Group
Texas Tech University
Lubbock, Texas**

ABSTRACT

This research focuses on the role of radiation-dynamics interaction in regional numerical weather prediction of severe storm environment and mesoscale convective systems over the central United States. Based upon the earlier numerical model simulation experiments, we believe that such interaction can have profound impact on the dynamics and thermodynamics of regional weather systems. The research will be carried out using real-data model forecast experiments performed on the CRAY-X/MP computer at NASA/MSFC.

The forecasting system to be used in this study is a comprehensive mesoscale prediction system which includes analysis and initialization, the dynamic model, and the post-forecast diagnosis codes. The model physics are currently undergoing many improvements in parameterizing radiation processes in the model atmosphere.

The forecast experiments in conjunction with in-depth model verification and diagnosis are aimed at a quantitative understanding of the interaction between atmospheric radiation and regional dynamical processes in mesoscale models as well as in nature. Thus, significant advances in regional numerical weather prediction can be made. Results shall also provide valuable information for observational designs in the area of remote-sensing techniques to study the characteristics of air-land thermal interaction and moist processes under various atmospheric conditions.

ACKNOWLEDGEMENTS

I am grateful for the opportunity to participate in the Summer Faculty Fellowship Program at NASA/MSFC. I would like to express my appreciation to Dr. Mike Freeman and Ms. Ernestine Cothran (CO-Directors) and their staff for very professional management of the program. Also, I like to thank the Department of Mechanical Engineering at UAH for providing a summer faculty ID which allows me to use the University facilities.

I would like to thank my colleagues Dr. F. Robertson of NASA/MSFC and L. M. Kolb of USRA for their interest in this research and providing valuable discussions and comments on this work. I wish to thank Mr. M. Seabloom, Ms. J. Srikishen, and Mr. M. Smith for technical help in carrying out numerical experiments on the CRAY-X/MP computer at MSFC.

I. INTRODUCTION

One of the central issues in regional numerical weather prediction (NWP) is the forecasting of mesoscale systems driven by latent heating because of their often severe nature. The present study will focus on the role of radiation-dynamics interaction in the numerical forecast of mesoscale convective storms of horizontal scale ranging from 50 km to 500 km in the central United States during severe weather seasons of spring and early summer.

It is well known that radiation is the primary energy source for the generation and maintenance of the dynamic systems in the atmosphere. The dynamic systems develop within baroclinic patterns as a result of horizontal differences in radiation budgets. On the other hand, the cloud and moisture distribution associated with the dynamic systems will modify the radiation field.

However, in short-range (<36 hours) regional NWP, the atmospheric radiation is generally regarded as a minor diabatic process because the rate of radiative warming/cooling in the troposphere is quite small in comparison with that of latent heating. The mean rate of net temperature change in the atmosphere under the clear sky condition due to the absorption of insolation and the emission of infrared irradiance by air is on the order of -1°C per day. To date, an explicit treatment of radiation processes in the model atmosphere does not exist in any comprehensive regional models (Pielke, 1984). And there is still lack of any in-depth study of the role of radiation-dynamics interaction in the short-range numerical forecasting of severe convective storms.

Nevertheless, based on earlier case studies we feel that the radiation-dynamics interaction may exert profound impact on the development of mesoscale storms over the central United States from the standpoint of the following three aspects.

- A. Effects of cloudiness associated with mesoscale systems on regional radiation budget.
- B. Surface heating and moisture source.
- C. Synoptic instability in the planetary boundary layer (PBL).

These three aspects will be discussed in some detail in Section II.

Scientific questions concerning the impact will be addressed using real-data model simulation experiments followed by detailed model verification and diagnosis. The prediction system to be used in this study is the Drexel LAMPS (Limited Area and Mesoscale Prediction System) which is a comprehensive mesoscale prediction system. LAMPS includes analysis and initialization, the dynamic model, and the post-forecast diagnosis codes. A detailed description of LAMPS can be found in the papers by Perkey (1976), Kreitzberg (1978), and Chang *et al.* (1981).

The model initial state is obtained from the objective analysis of rawinsonde observations based upon the isentropic scheme designed by Bleck (1975). The dynamic model is a three-dimensional primitive equation model consisting of 15 terrain following vertical levels and a longitude-latitude horizontal grid. The model levels ranging from 0 to 16 km with relatively higher resolution in the PBL. The model physics consist of essential dry and moist processes for the quantitative study of various aspects of convective storms under diverse meteorological conditions. A brief discussion of solving the radiative transfer equations in the model is included in Section III. The numerical experiments and their purposes, and related hypotheses to be tested are described in Section IV.

II. RADIATION-DYNAMICS INTERACTION

A. The Effects of Cloudiness on Radiation Budgets

Clouds are the principal modulators of radiation in the atmosphere. The presence of clouds increases the counter radiation and, hence, decreases the effective outgoing infrared radiation and incoming solar radiation. The infrared cooling near the cloud top can be as high as 5°C per day (Katayama, 1966; Stephens and Webster, 1981) which is much greater than the average cooling rate under the clear sky conditions, while for the layers below the cloud the cooling rates are considerably reduced. Thus, the distribution of horizontal-developing clouds associated with dynamic systems can have strong influences on the vertical temperature profile throughout the troposphere and the regional baroclinic patterns (cloudy vs. clear sky areas).

The significance of the strong cloud-top cooling may be illustrated by some results from regional model simulations of a developing cyclone. Figure 1 shows the 24-h forecast height, temperature and vector wind differences on the 700, 500, and 300 mb pressure surfaces, respectively, between the wet (with latent heat release) and dry simulations for the 20 May 1977 SESAME case (Chang *et al.*, 1982). The maximum differences occurred over the regions of intense precipitation. The model used in the simulations did not include the effects of strong cloud top cooling.

It is conceivable that the temperature differences at 300 mb in the heavy rain areas over Nebraska and Kansas would have been reduced by a few degrees if the strong cloud top cooling had been considered in the wet model. This could greatly modify the upper air circulations and the development of the surface cyclone.

The substantial reflection of solar radiation takes place at the cloud top because of its large albedo which, except for cirrus, is on the order of 0.5 (Stephens and Webster, 1981). This depletion process may be insignificant so far as the absorption of insolation in the atmosphere is concerned. However, it will have controlling effects on surface energy budgets over the land where solar radiation, infrared radiation, and sensible heat flux represent three major heat exchange processes. The horizontal differences in surface energy budgets, again cloudy vs. clear sky conditions, will contribute to the low-level baroclinicities.

B. Surface Heating and Moisture Source

Because the earth's surface receives more radiative energy from the sun than it loses in the form of infrared radiation during the daytime, some of the excess in radiative energy will be transferred from the ground to the atmosphere in the form of sensible heat flux. The intensity of sensible heating is largely dependent upon the characteristics of the earth's surface and the atmospheric conditions. In arid regions, such as the subtropical deserts, surface heating plays a vital role in the low-level circulations (Chang, 1980; Tang and Reiter, 1984).

The moisture sources for the development of convective storm are closely correlated with the low-level circulations. In the central United States, a major source is the northward warm moist air streams

originating over the Gulf of Mexico. During the severe weather seasons of spring and early summer, one of the major mechanisms contributing significantly to the maintenance of such northward flows of warm moist air is believed to be surface heating over the arid regions of Mexico and the southwestern United States.

Particularly, over the Mexican highlands due to very low moisture content in the air and the underlying surface, and the near-vertical solar position at local noon solar radiation works very effectively in heating the earth's surface. The thermal low generated as a result of intense surface heating enhances the east-west pressure gradient across the Gulf of Mexico and consequently the northward flows of warm moist air. Hence, surface heating can have a direct impact on severe storm development in the central United States. The neglect of such an impact in regional NWP may result in an underprediction of convection in the model.

Figure 2 shows a) the simulated 24-h 500-mb height and wind speed, b) the corresponding verification map (NMC analyses), and c) the simulated 24-h 300-mb vorticity and 500-mb vertical motion for the severe storm case of 10-11 April 1979. The intensity and patterns of the model systems are in good agreement with the observations. The structure of the upper air jet and low suggests strong synoptic-scale forcing over Oklahoma and Texas where enduring severe convection was observed (Moore and Fuelberg, 1981). Despite these well-simulated dynamic systems, the model did not predict the longevity of the organized convections in the central United States.

The forecast deficiency was primarily attributed to the lack of moisture supply. Surface temperature verification maps (not shown) revealed systematically lower model surface temperatures over Mexico were noted. Consequently, the model failed to generate a surface low over the Mexican highlands and shifted the streams of warm moist air toward the Gulf States instead of Oklahoma and Texas as observed.

Also, Carlson *et al.* (1983) suggested that strong surface heating over the Mexican plateau in conjunction with synoptic-scale motion was responsible for the creation of an elevated mixed-layer inversion over Texas. The inversion prevents the outbreak of thunderstorms over a large area and restricts convection to regions of relatively weak stability.

Lanicci (1984) used numerical experiments to test the influences of soil moisture on the severe storm environment. He suggested that the diabatic effects of radiation processes at the land surface were strongly dependent on the soil moisture and its horizontal gradient, and significantly influenced the severe-storm environment including the dryline, low-level jet, latent instability, etc.

C. Dynamic Instability in the Planetary Boundary Layer

Based on linearized perturbation analysis in conjunction with the numerical experiments, Chang (1987) concluded that in summer over the tropical North Africa dynamic instability possessing the prominent characteristics of African wave disturbances can occur in the lower troposphere characterized by relatively marked meridional temperature gradient ($\sim 1^\circ\text{C}$ per 100 km) and near-neutral stability as in the unstable PBL. The deep near-neutral layer is created by strong surface heating due to solar radiation. As shown in Fig. 3, the instability results in the perturbations of maximum growth rate having a horizontal scale (around 2000 km) very close to the mean wavelength of African disturbances. Reduction in the horizontal thermal gradient shifted the most unstable mode toward the short waves.

Over the southwestern deserts of the United States, the short-wave disturbances of similar horizontal scale as African waves are often observed in the lower troposphere in summer under relatively weak synoptic (undisturbed) conditions. Upon encountering a potentially unstable tropical air mass, the disturbances can trigger heavy convective rainfall. The disturbances appear to originate over the southwestern deserts with a deep PBL. The environmental conditions of the southwestern deserts resemble those of North Africa in summer. The short waves may have an origin similar to African waves. If so, an adequate treatment of radiation processes in regional-scale models is essential for improving some flash flood forecasting.

III. RADIATIVE TRANSFER

The physical processes that differ from those of the LAMPS model used in previous case studies are the treatment of radiative transfers and surface energy balance. In modeling radiative transfers and the vertical

heat exchanges over land, the usual constraints of a numerical model, i.e., not to be too time-consuming, must be considered. For example, an exact treatment by integrating the radiative transfer equations over wavelength and optical path in the model atmosphere is not desirable at this time. Some simplifications and approximations as described below are made in modeling these diabatic processes.

To solve the transfer equations for infrared radiation a simplified method based on an emissivity technique will be used. The capability of the emissivity technique has been studied by many researchers (Rodgers, 1967; Fels and Schwarzkopf, 1975; Chang, 1980). The results indicated that the technique is an acceptable method for numerical prediction models. The differences in computed cooling rates between the emissivity approximation and the Goody (1964) random model are on the order of 0.1°C per day under a wide variety of atmospheric conditions (Fels and Schwarzkopf, 1975). For modeling solar radiation, a technique designed for the UCLA general circulation model (Haltiner and Williams, 1980) and also used in the FSU tropical prediction model (Chang, 1980) will be adopted for this study. Only water vapor is considered as an optically active gas in the model atmosphere.

The energy balance at the land surface involves solar and infrared irradiance, and sensible and latent heat flux, assuming zero soil heat capacity. The assumption is anticipated to have little impact on the nature of surface heating in short-term NWP. Surface heating is well known to be a strong function of cloud cover and moisture content in air. The large change of soil temperature in response to radiation occurs in a layer of few cm thick in a 12-h period (Sellers, 1967). The specific heat capacity of most soils is on the order of $0.3 \text{ cal gm}^{-1} \text{ K}^{-1}$. The energy required for a drastic warming rate, for example 3°C per hour, in the thin layer is less than 5 percent of the incoming irradiance or the outgoing heat flux. Also, heat conduction inside the subsurface layer is neglected.

The sensible and latent heat fluxes are related by a ground wetness parameter which is approximated as a linear function of saturation ratio predicted in the model surface layer. The parameter resembling the Bowen ratio ranges from 2 over the dry desert to 0.5 over the wet surface. This energy balance approach in determining heat flux over land appeared to

produce quite realistic results in the model simulations of the low-level circulations over North Africa by Chang (1980).

A unique feature of LAMPS is the explicit forecast of cloud water which is treated as an important component of moisture conservation in the model atmosphere. In comparison with the parameterization approach, this represents a more dynamically consistent way in determining the location of model cloud layers. Thus, the influences of cloudiness on radiation field and surface energy balance, and the resultant effects on the dynamical systems can be better resolved.

IV. OBJECTIVES

The purpose of the research is to obtain a quantitative assessment of the impact of radiation-dynamics interaction on the regional NWP of mesoscale storms. For this purpose, we will carry out in-depth model forecast experiments on the severe weather events occurring on 2 April 1982. The experiments to be conducted on the NASA CRAY-X/MP computer for the case study and the related scientific questions to be stressed are summarized below.

- 1) A 24-h fine-mesh forecast with the complete model physics.
- 2) Similar to 1), except without the effects of cloudiness in radiative warming/cooling computation.
- 3) Similar to 1), except without strong surface heating over arid land.

The fine-mesh model ($\Delta x \approx 140$ km) is intended for creating the severe storm environment. The first experiment is also regarded as a control run. The simulated results will be compared with observations for evaluating the model performance. Experiment 2 is designed to test the hypothesis that the effects of cloudiness on radiative temperature changes over the areas of heavy precipitation will cause significant dynamic responses and thus alter the storm's structure. Experiment 3 is designed to test the hypothesis that surface heating over the arid land, for example the Mexican highlands, can affect the moisture availability for storm development in the south-central United States. The case study chosen represents a strongly disturbed situation. Thus, the question concerning the PBL dynamic instability as mentioned in II.c may not be addressed with this case study.

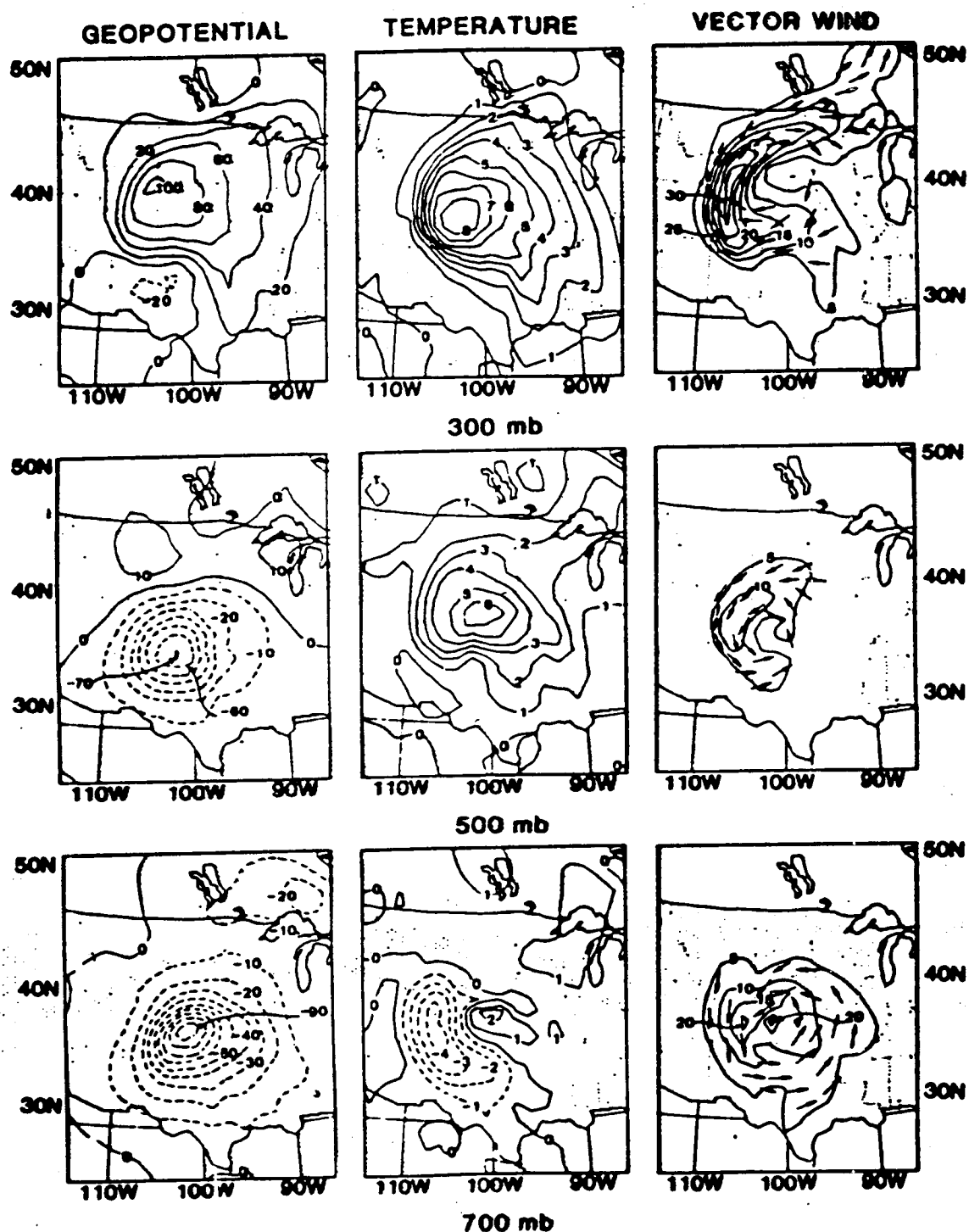
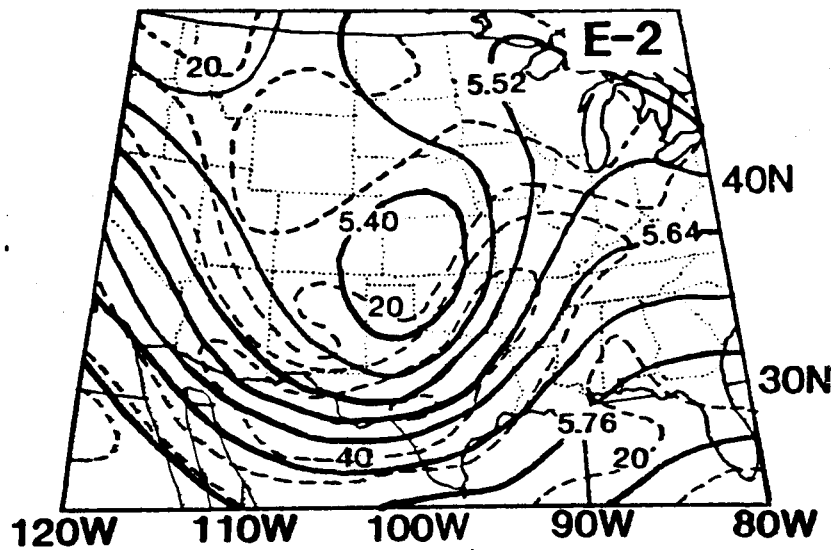
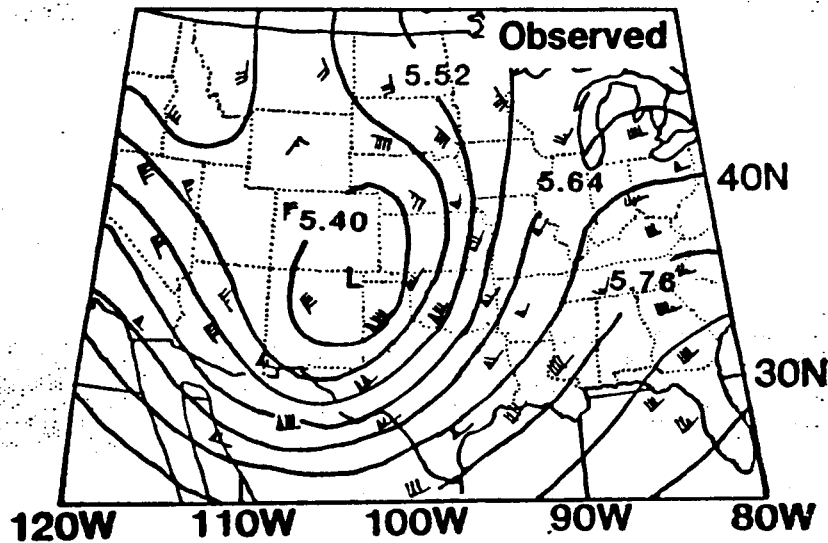


Figure 1. Simulated 24-h difference fields (wet - dry run) at 1200 GMT 21 May 1977. The geopotential height, interval is 10 m on both 700- and 500-mb levels and 20 m on the 300-mb level, the temperature interval is 1°C and the isotach interval is 5 m s⁻¹ for all three levels. The dashed contours denote negative values.

ORIGINAL PAGE IS
OF POOR QUALITY



(a)



(b)

Figure 2. (a) Simulated 24-h 500-mb height (solid lines in km) and wind speed (dashed lines in m s^{-1}) at 1200 GMT 11 April 1979, and (b) observed conditions.

V. RESULTS

Some preliminary results of the proposed numerical model experiments performed on 2 April 1982 case are presented. We find that although the atmospheric radiation has shown consistently cooling effects, the mean temperature of the model atmosphere has actually increased due to the increase in latent heating in this case. Figures 3a and 4a show, respectively, the simulated 12-h and 24-h accumulation of convective precipitation for the reference run. The corresponding results for the model with the improved radiative transfer computation including the effects of cloudiness are presented in Figs. 3b and 4b. The differences between the two simulations are quite striking.

The 12-h simulations indicate that the improved radiation computation increases the maximum precipitation by about half inch in southeastern Missouri and generates a center of convection off the coast of Florida (Fig. 3b). For the 24-h simulations, Fig. 4a shows one major center with intense convection confined to the central mid-west states and a relatively weak maximum off the coast, while Fig. 4b shows multi-center structure with wide-spread convection over the nation. The former has no convective but substantial non-convective rainfalls (not shown) over Montana. Unfortunately, at this time we do not have adequate precipitation information for a quantitative model verification.

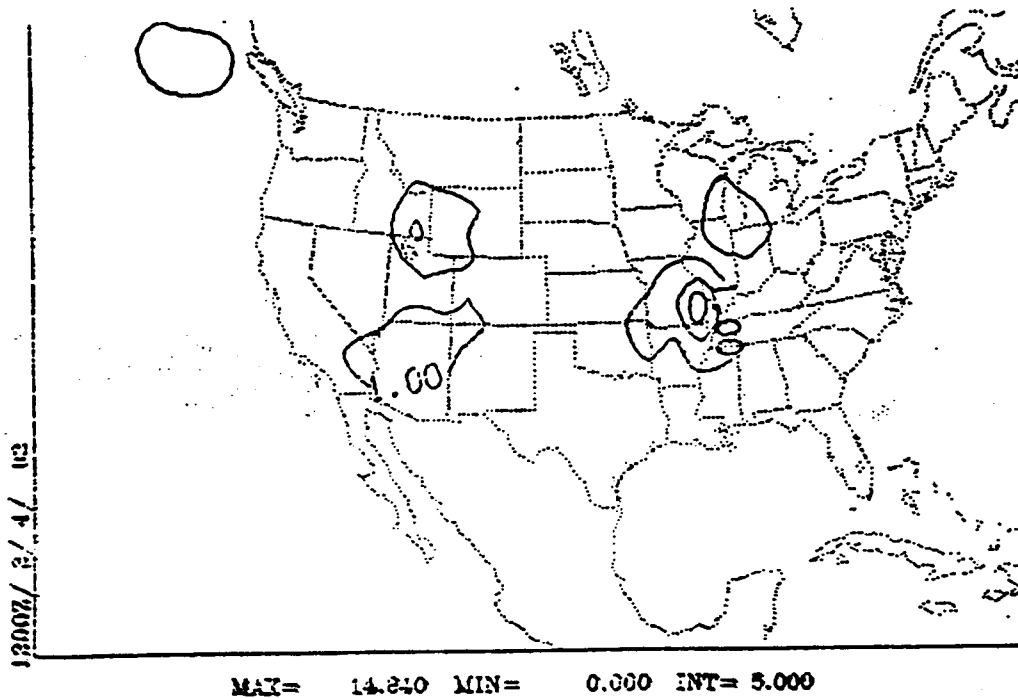
VI. CONCLUSIONS AND RECOMMENDATIONS

The impact of radiation-dynamics interaction on the regional model simulations is obvious. It appears that radiative transfer results in the destabilization of the model atmosphere. The model convective activities are enhanced and the model spin-up processes may be altered.

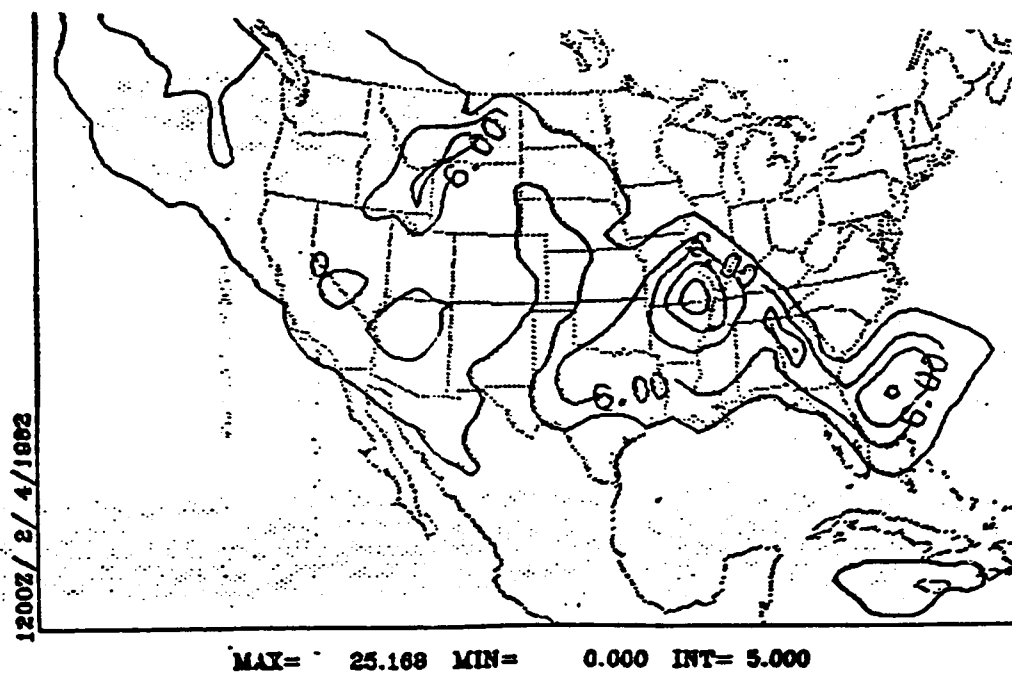
We will carry out more numerical experiments in addition to those proposed earlier and engage in the in-depth diagnoses of model output. This will include other case studies and followed by energy diagnosis, the analysis of relevant thermodynamic and dynamic variables, and the boundary layer convergence of moisture. After some quantitative assessment, we shall next investigate the radiation-dynamics interaction in a finer mesh model (e.g., 70 km grid size).

It is believed that regional models with a realistic treatment of atmospheric radiation provide a tool for testing the usefulness of satellite observations in studying moist processes in the atmosphere. It is also desirable to compare, if possible, the vertical profiles of model-generated with those of satellite-observed irradiance for further improvement in parameterizing model radiative transfer.

The ultimate goal of this effort is to achieve a better understanding of the diabatic effects of radiation on the regional weather systems and to improve regional-scale NWP.



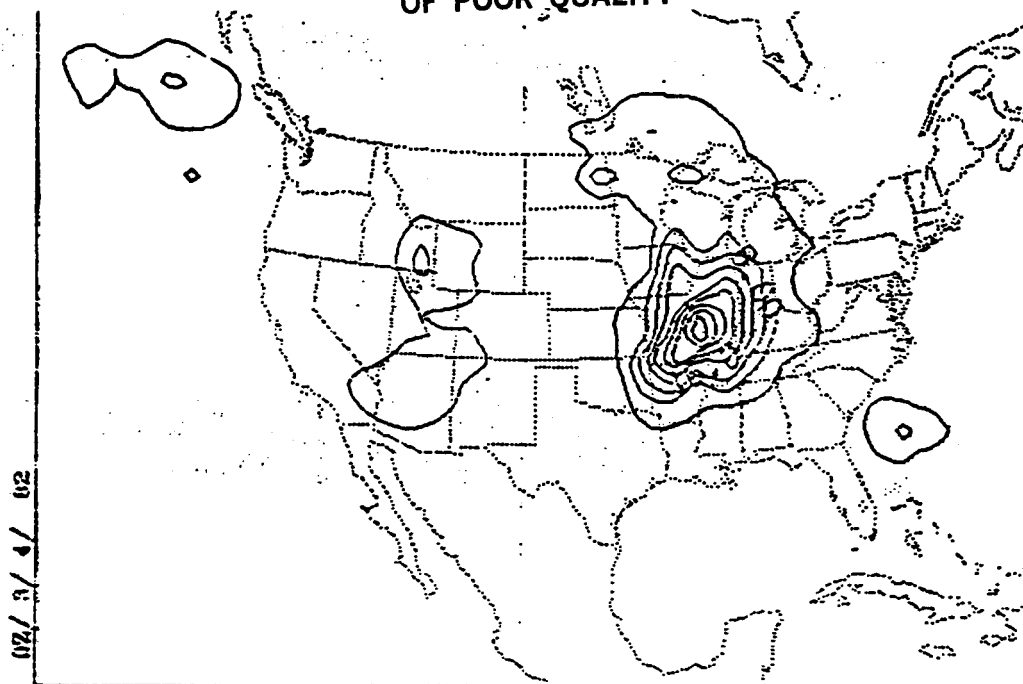
(a)



(b)

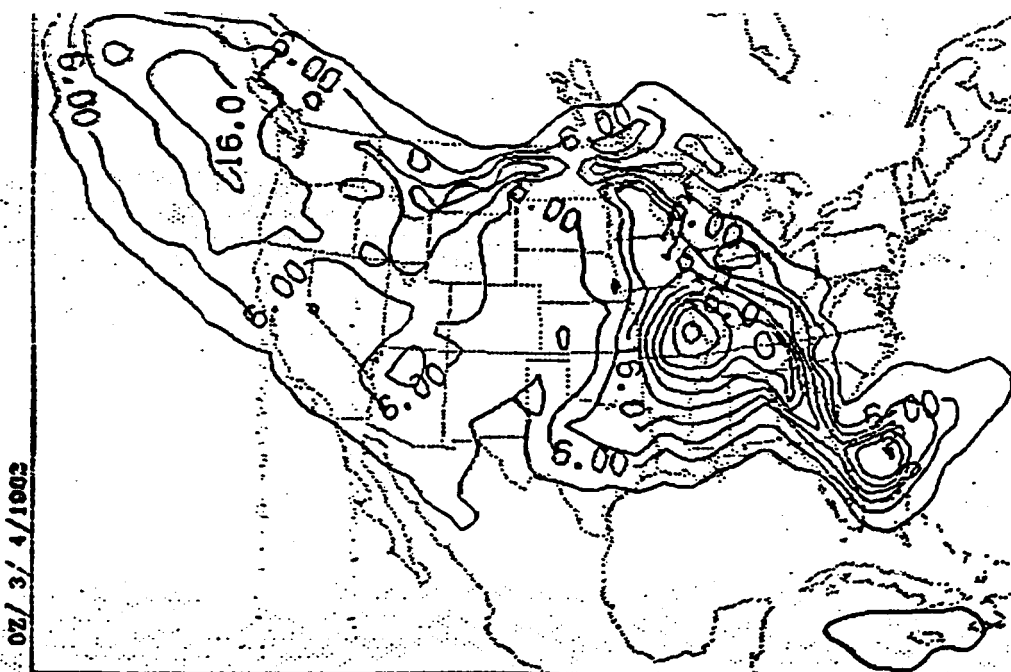
Figure 3. Simulated 12-h accumulated convective precipitation for the case of 2 April 1982. (a) without and (b) with improved radiative transfer. Contour interval is 5 mm.

ORIGINAL PAGE IS
OF POOR QUALITY



MAX= 45.027 MIN= 0.000 INT= 5.000

(a)



MAX= 49.327 MIN= 0.000 INT= 5.000

(b)

Figure 4. Similar to Fig. 3 except for the 24-h simulations.

VII. REFERENCES

- Bleck, R., 1975: An economical approach to the use of wind data in the optimum interpolation of geo- and Montgomery potential fields. *Mon. Wea. Rev.*, **104**, 807-816.
- Carlson, T. N., S. G. Benjamin, G. S. Forbes, and Y.-F. Li, 1983: Elevated mixed layers in the regional severe storm environment. *Mon. Wea. Rev.*, **111**, 1453-1473.
- Chang, C.-B., 1980: On the influences of solar radiation and diurnal variation of surface temperature on African disturbances. Ph.D dissertation, The Florida State University, 157 pp.
- Chang, C.-B., 1987: A theoretical and numerical study of African wave disturbances. *J. Meteor. Atmos. Phys.*, **37**, 159-170.
- Chang, C. B., D. J. Perkey, and C. W. Kreitzberg, 1981: A numerical case study of the squall line of 6 May 1975. *J. Atmos. Sci.*, **38**, 1601-1615.
- _____, _____, and _____, 1982: A numerical case study of the effects of latent heating on a developing wave cyclone. *J. Atmos. Sci.*, **39**, 1555-1570.
- Fels, S. B., and M. D. Schwarzkopf, 1975: The simplified exchange approximation: A new method for radiative transfer calculation. *J. Atmos. Sci.*, **32**, 1475-1488.
- Goody, R. M., 1964: Atmospheric Radiation. *Oxford University Press*, 611 pp.
- Haltiner, G.J., and R. T. Williams, 1980: Numerical Prediction and Dynamic Meteorology, 2nd edition. *Wiley, New York*.
- Katayama, A., 1966: On the radiation budget of the troposphere over the northern hemisphere (I). *J. Meteor. Soc. Japan*, **44**, 381-401.
- Kreitzberg, C. W., 1978: Progress and problems in regional numerical weather prediction. *Proc. SIAM*, Vol. 11, Amer. Math. Soc., 32-58.

- Lanicci, J. M., 1984: The influences of soil moisture distribution on the severe-storm environment of the southern Great Plains: A numerical study of the SESAME IV case. M.S. thesis, The Pennsylvania State University, 237 pp.
- Moore, J. T., and H. E. Fuelberg, 1981: A synoptic analysis of the first AVE-SESAME '79 period. *Bull. Amer. Meteor. Soc.*, **62**, 1577-1590.
- Perkey, D. J., 1976: A description and preliminary results from a fine-mesh model for forecasting quantitative precipitation. *Mon. Wea. Rev.*, **104**, 1513-1526.
- Pielke, R. A., 1984: Mesoscale Meteorological Modeling. *Academic Press*, 612 pp.
- Rodgers, C. D., 1967: The use of emissivity in atmospheric radiation calculation. *Quart. J. Roy. Meteor. Soc.*, **93**, 43-54.
- Sellers, W. D., 1967: Physical Climatology. *The University of Chicago Press*, 272 pp.
- Stephens, G. L., and P. J. Webster, 1981: Clouds and climate: Sensitivity of simple systems. *J. Atmos. Sci.*, **38**, 235-247.
- Tang, M., and E. R. Reiter, 1984: Plateau monsoon of the northern hemisphere: A comparison between North America and Tibet. *Mon. Wea. Rev.*, **112**, 617-637.

N89-21733

1988 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER

THE UNIVERSITY OF ALABAMA

COMPUTER CODES FOR
THERMAL ANALYSIS OF A SOLID ROCKET MOTOR NOZZLE

PREPARED BY:	Rajinder Singh Chauhan
ACADEMIC RANK:	Assistant Professor
UNIVERSITY AND DEPARTMENT:	North Carolina A&T State University/Mechanical Engineering
NASA/MSFC:	
LABORATORY:	Structures and Dynamics
DIVISION:	Thermal Engineering and Life Support
BRANCH:	Thermal Systems Branch
MSFC COLLEAGUE:	Kenneth McCoy
DATE:	August 8, 1988
CONTRACT NO:	NGT 01-002-099 The University of Alabama

COMPUTER CODES FOR
THERMAL ANALYSIS OF A SOLID ROCKET MOTOR NOZZLE

by

Rajinder Singh Chauhan

N.C.A. & T. State University

Greensboro, North Carolina

ABSTRACT

A number of computer codes are available for performing thermal analysis of solid rocket motor nozzles. Aerotherm Chemical Equilibrium (ACE) computer program can be used to perform one-dimensional gas expansion to determine the state of the gas at each location of a nozzle. The ACE outputs can be used as input to a computer program called Momentum/Energy Integral Technique (MEIT) for predicting boundary layer development, shear and heating on the surface of the nozzle. The output from MEIT can be used as input to another computer program called Aerotherm Charring Material Thermal Response and Ablation Program (CMA). This program is used to calculate ablation or decomposition response of the nozzle material. A new code called Failure Analysis Nonlinear Thermal and Structural Integrated Code (FANTASTIC) is also likely to be used for performing thermal analysis of solid rocket motor nozzles after the program has been duly verified. A part of the verification work on FANTASTIC was done by this author in the summer of 1987 by using one and two-dimension heat transfer examples with known answers. During this summer (1988), an attempt has been made to prepare input for performing thermal analysis of CCT nozzle using FANTASTIC computer code. The CCT nozzle problem will first be solved by using ACE, MEIT, and CMA. The same problem will then be solved using FANTASTIC. These results will then be compared for verification of FANTASTIC.

ACKNOWLEDGEMENTS

I wish to acknowledge the NASA/ASEE Summer Faculty Fellowship Program, along with Mike Freeman, 1988 University of Alabama Co-Director and Ernestine Cothran, MSFC University Affairs Officer.

I would like to Thank Dick Wegrich, Chief of Thermal Control Engineering Branch for providing office space and work related facilities.

I owe my gratitude to NASA/MSFC colleague Kenneth McCoy for having me work with him. His guidance and help provided valuable insight in carrying out my work at MSFC.

I also would like to thank Dallas Clark, Thermal Analysis Branch for his support in my work.

I am grateful to Betty K. Golden of Thermal Control Engineering Branch for typing this manuscript.

INTRODUCTION

Failure Analysis Associates (FAA), Palo Alto, California has developed an advanced computer code for MSFC to improve the accuracy of the solid rocket nozzle motor analysis. The computer code is called "FANTASTIC", an acronym for Failure Analysis Nonlinear Thermal And Structural Integrated Code. This code incorporates a higher order integration scheme for modeling steep temperature and strain gradients. It has a modular design and is intended as a multi-purpose thermostructural analysis code.

The first version of the above code was released by FAA in the beginning of 1987. This author worked as NASA/ASEE Summer Faculty Fellow at MSFC in the summer of 1987 and got an opportunity to study and verify this code. A number of one and two-dimensional heat transfer problems were developed and run on FANTASTIC for trial. The problems faced during these trial runs were reported to the developer. As a result, the code was improved and a second revised version of the program was released which could solve simple problems.

In the summer of 1988, the author returned to MSFC as a Summer Faculty Fellow and started working on FANTASTIC but with a different approach. The aim was to verify FANTASTIC for the analysis of solid rocket motor nozzles (SRM). This prompted the author to design a problem which could be first solved using existing codes with MSFC.

CCT-5 nozzle was selected as an example. The author studied one existing code called Aerotherm Chemical Equilibrium (ACE) computer program. Other codes which will be needed to make a complete thermal analysis for CCT-5 nozzle are called Momentum/Energy Integral Technique (MEIT) and Aerotherm Charing Material Thermal Response and Ablation Program (CMA). A brief review of running ACE and pertinent information about these codes is given in the next section.

ACE COMPUTER PROGRAM

ACE program was developed by Acurex Corporation, Mountain View, California for MSFC. It is an extremely versatile code for calculations of a number of thermochemical processes taking place in closed or open system. A few types of calculations which may be performed by ACE are:

1. The thermodynamic and chemical state in the combustion chamber of the SRM may be calculated by specifying the elemental composition of the injected fuel.
2. ACE can perform an isentropic expansion of the combustion gases and boundary layer edge state tables may be calculated for subsequent input to heat conduction and ablation energy balance programs.
3. The thermodynamic state at the surface of an ablation material may be calculated as a function of pyrolysis gas and char rates normalized with respect to mass transfer coefficients.

Combustion Chamber Gas Expansion Calculation Using ACE

The input data card deck for running ACE may consist of the following seven card sets:

1. Control card
2. Surface fluxes
3. Oblique or normal shock data
4. Elemental composition data
5. Fuel temperature and diffusion factor data
6. Species thermochemical equilibrium data
7. Reaction rate data

Card set 1 is a control card which contains a 10 element array, KR(I). This array controls most of the program options and tells the program what to expect from the remaining card sets. Few problems require all card sets.

As an example the input for performing combustion chamber calculation and further gas expansion for a range of pressure values is given below:

Card Set 1

Control Card (1 Card)

Field 1

(Column 1-10) this is the variable array KR(I) which is used to control the various program options

Column 1: State Option

0 for "Assigned temperature"

Column 2: Elemental Composition

1 for "Input new elemental compositions" in card set 4.

Column 3: Species thermochemical data

1 for "Read from cards new species thermochemistry data (card set 6) and, if provided, data on fail temperatures and diffusion factors (card set 5)

Column 4: Mass balance options

0 for "Closed system mass balance"

Column 5: Expansion Option

2 for "Save the current enthalpy as a chamber enthalpy, and save the current entropy, for subsequent isentropic expansion calculations.

Columns 6 through 10 "0"

Field 2

(Columns 11-20) The thermodynamic state variable Y since in column 1 of Field 1, KR(1) = 0, therefore, Y is assigned to temperature ($^{\circ}$ k).

(Y = 3430.0 for this example)

Field 3

(Columns 21-30) Pressure variable, PR
If PR 0, PR is assigned to the system pressure in atmospheres.
(PR = 68.0735 for this example).

Field 4 - Field 6

(Columns 31-80)

No input needed for this example

Card Sets 2-3

No input needed

Card Set 4

Card 1

(Columns 1-3) The number of elements in the system

Cards 2, 3, 4,(one such card for each element)

Field 1

(Columns 1-3) the atomic number of the element

Field 2

(Columns 4-15) the name of the element

Field 3

(Columns 16-25) the atomic weight of the element

Field 4

(Columns 26-55) the relative amounts of the element in components 1, 2, and 3, respectively.

The input for the current example for Card Set 4 is given below:

COLUMNS

	<u>(1-3)</u>	<u>(4-15)</u>	<u>(16-25)</u>	<u>(26-55)</u>
Card 1	6			
Card 2	1	Hydrogen	1.008	3.7176
CARD 3	6	Carbon	12.011	0.9407
Card 4	7	Nitrogen	14.011	0.6263
Card 5	8	Oxygen	16.000	2.4396
Card 6	13	Aluminum	26.982	0.5930
Card 7	17	Chlorine	35.457	0.5952

CARD SET 6

SPECIES THERMOCHEMICAL EQUILIBRIUM DATA

(Number of cards = $1 + 3 \times$ number of species) see note

If KR(3) = 0,5,6,7,8 or 9, skip this card set

There are three of these cards for each molecular, atomic, condensed, ionic, or electron species. The end of this card set is signaled by a blank card (see Reference 3 for additional discussion relative to this card set)

CARDS 1,4,7 . . . Describe the elemental composition of the species and establish its name designation

FIELDS 1,3,5,...,13. (one for each element in the species)

(Columns 1-3, 7-9, 13-15,...,37-39, each format F3.0)

The number of atoms (of atomic number given in the following field) in a molecule of this species

(If field 1 is zero, this card is presumed to represent the end of Card Set 6)

FIELDS 2,4,6,...,14 (one for each element in the species)

(Columns 4-6,10-12,16-18,...,40-42, each format I3)

The atomic numbers of the elements in the molecule (the number of atoms of which was given in the previous field)

(If field 2 is zero, this card is presumed to be the first card of Card Set 5)

Ionized species are described by the addition or subtraction of an electron (atomic number 99). For example, NO^+ would be described in fields 1 through 6 as bb1bb7bb1bb8b-1b99.

FIELD 15

(Columns 43-72, Format 7A4A2) the source and date of the thermochemical data for this specie. Used for output only

FIELD 16

(Columns 73-80, Format 2A4) the name designation of this species (e.g., AL203). This variable is used for output and as a means of identifying data entered in Card Set 5

FIELD 16 (Continued)

NOTE: The ACE code is dimensioned such that the number of chemical species may not exceed 149. Also the number of condensed phase species may not exceed (14-number of elements).

CARDS 2,5,8...Lower temperature range thermochemical data

FIELDS 1,2,3,4,5 and 6

(Columns 1-54, Format 6E9.6) Input the 6 constants (F1,F2,F3,F4,F5,F6) appropriate to the lower temperature range of the thermodynamic data for this species. These constants are defined as follows, where T is in °K:

F1 = the heat of formation of the species at 298°K from the JANAF base state (elements in most natural form at 298°K) in cal/mole

F2 = the enthalpy change of the species from 298°K to 3000°K in cal/mole

F3,F4 and F5 are defined by a curve fit to the heat capacity at constant pressure of the form:

$$C_p = F3 + F4 T + \frac{F5}{T^2} \text{ in cal/mole } ^\circ\text{K}$$

F6 = the entropy of the species at 3000°K in cal/mole °K

FIELD 7

(Columns 61-66, Format F6.0) the upper limit of the lower temperature range in °K

FIELD 8

(Column 67, Format II) the phase specification:

- 1 signifies gaseous species
- 2 signifies solid species
- 3 signifies liquid species

The only phase combination allowed in one three card set is solid-liquid in which case a 2 and 3 would appear in fields 8 of cards 2 and 3, respectively.

CARDS 3,6,9...Upper temperature range thermochemical data

These cards are the same as cards 2,5,8...except use constants for the upper temperature range and field 7 is ignored

LAST CARD OF CARD SET 6 MUST BE BLANK

OVERALL PROCEDURE OUTLINE AND CONCLUSIONS

An outline of the procedure to be following for analysing the SRM nozzle problem using different compuer codes is as follows:

1. Preparation of input for ACE
2. Run ACE expansion
3. To arrange the output from ACE in an acceptable format for running MEIT
4. Run MEIT for boundary layer analysis
5. Preparation of ablation tables using outut from MEIT.
The ablation tables are used as input for CMA
6. Run CMA
7. Work on input for FANTASTIC
8. Run FANTASTIC
9. Compare the results from CMA with those from FANTASTIC

The first three steps have been completed. The work on last the last three steps can be started after FANTASTIC has been revised for handling nozzle problem. The work on steps 4-6 can be continued.

ORIGINAL PAGE IS
OF POOR QUALITY

N89-21734

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

RADIATION EFFECT ON ROCKET ENGINE PERFORMANCE

Prepared by:	Huei-huang Chiu, Ph.D.
Academic Rank:	Professor
University and Department:	University of Illinois at Chicago Department of Mechanical Engineering
NASA/MSFC	
Laboratory:	Propulsion
Division:	Propulsion Systems
Branch:	Performance Analysis
MSFC Colleague:	Klaus W. Gross
Date:	September 1988
Contract No.:	NGT-01-022-099 The University of Alabama

RADIATION EFFECT ON ROCKET ENGINE PERFORMANCE

by

Huei-huang Chiu

Professor of Mechanical Engineering
University of Illinois at Chicago
Chicago, Illinois

ABSTRACT

The effects of radiation on the performance of modern rocket propulsion systems operating at high pressure and temperature have been recognized as a key issue in the design and operation of various liquid rocket engines of the current and future generations.

Critical problem areas of radiation coupled with combustion of bipropellants are assessed and accounted for in the formulation of a universal scaling law incorporated with a radiation-enhanced vaporization combustion model. Numerical algorithms are developed and the pertaining data of the Variable Thrust Engine (VTE) and Space Shuttle Main Engine (SSME) are used to conduct parametric sensitivity studies to predict the principal intercoupling effects of radiation. The analysis reveals that low enthalpy engines, such as the VTE, are vulnerable to a substantial performance set back by the radiative loss, whereas the performance of high enthalpy engines such as the SSME, are hardly affected over a broad range of engine operation. Additionally, combustion enhancement by the radiative heating of the propellant has a significant impact in those propellants with high absorptivity. Finally, the areas of research related with radiation phenomena in bipropellant engines are identified.

ACKNOWLEDGMENTS

I would like to extend my great appreciation to Mr. Klaus W. Gross, MSFC, who suggested these research topics and rendered valuable technical and scientific inputs on broad basic and practical issues in modern liquid rocketry. This collaborative effort has significantly broadened the scope and the depth of this preliminary research. Computational codes have been developed jointly by Mr. Gross and Mr. A. N. Krebsbach, and executed by Mr. Krebsbach. Their contributions form the base of the principal conclusion and are greatly appreciated. Special thanks are also due to Mr. Gross for his careful review of the manuscript and helpful suggestions. I wish to express my appreciation to the staffs of the Performance Analysis Branch, headed by Mr. R. Bachtel, for their hospitality and generous support. The excellent opportunity and active research environment provided by NASA, MSFC, and ASEE during the summer program under the skillful management of Ms. E. Cothran and Professor M. Freeman are graciously acknowledged.

1. INTRODUCTION

The performance of modern rocket engines (Sutton and Ross, 1975) is primarily affected by the competitive rate processes involving multi-phase turbulent reacting flow with phase change and heat transfer at an elevated chamber pressure, typically at 10^3 psia and chamber temperatures of 5500 to 6500°R at which levels the flame emission (Afgan et al. 1974) due to the molecular gas radiation plays two significant roles. Firstly, the radiation emission leaving the system constitutes a thermal energy loss and, thus, directly and indirectly affects the major performance parameters through the chamber temperature. These performance parameters include critical mass flux, specific impulse C^* , and total thrust. Secondly, the radiative heating of propellants enhances the vaporization and combustion of droplets and, thus, the process serves to improve the overall combustion efficiency of the motors.

Because of these two diametrically opposite roles played by radiation, the overall impact of flame emission may result in performance loss or gain depending on the relative preponderance of heat loss relative to propellant gasification rate enhancement. This constitutes the focal point of quantitative scrutinization.

The principal factors affecting the nature and extent of the impacts of radiation are traced down to the fundamental physics of emission, absorption, and scattering of gas molecules, droplets, soot particles in hydrocarbon propellants, and the combined radiation and conduction-convection heat transfer in the chamber (Summerfield, 1960). The practical significance of the above-mentioned factors depends on the basic design and operating parameters including propellant properties, mass flow rates, mixture ratio, chamber pressure, spray characteristics, and the combustor size.

One of the critical design issues of the current and future rocket engines is the lack of a comprehensive criterion that predicts and classifies the motor performance characteristics, in the presence of the radiation processes, in terms of the design and operating parameters described above.

Review of the literature accumulated over the past several decades reveals that there has been little or no in-depth research which pertains to the fundamental issues of radiation in liquid rocket combustion chambers. Furthermore, there has been a genuine lack of vital engineering data, modern analytical tools, and computational fluid dynamics (CFD) required for the analysis of the bipropellant and monopropellant combustion processes. In fact, many of the existing data and analytical tools developed for the radiative treatment of large scale industrial combustion appliances (Afgan et al., 1974) and air breathing engines are incomplete and inadequate for application in rocket motors for performance and design optimization. To this end, a comprehensive

motor performance scaling law is developed, on the basis of rudimentary thermodynamic and combustion principles, to aid in assessment and identification of the basic issues and qualitative prediction of the motor performance characteristics of modern liquid rocket engines.

The objectives of this study are (1) to develop an algebraic scaling law of motor performance and propellant droplet gasification laws, (2) to assess and compare the performance of the VTE and SSME with respect to regenerative heat recovery systems in each type of engine, and (3) to identify specific technical problem areas where radiation modeling and numerical simulation of radiation-coupled combustion processes are required.

2. SCALING LAWS OF ROCKET MOTORS

A performance scaling law is developed with (1) thermodynamic relationships that determine the chamber temperature as a function of propellant properties, chamber inlet conditions, wall heat transfer by conduction, convection, and radiation, combustion and vaporization efficiencies, and a regenerative efficiency; and (2) propellant droplet laws that predict the rate of vaporization and combustion under the effects of combined radiation and conductive heat transfer.

Such a law must preserve the universality for any type of propellants and engine configuration, regenerative or non-regenerative, and all engine power levels.

2.1 Thermodynamic Performance Model

A schematic, Figure 1 (a and b), shows a regenerative rocket engine that admits fuel and oxidizer at flow rate \dot{m}_F and \dot{m}_O , respectively, at an effective inlet temperature T_i . The reheater receives heat at the rate of q_R , radiative heat, and q_C , conductive heat, from the bulk of the gas in the combustion chamber. The fraction of heat received, η_n ($q_R + q_C$), is used in propellant preheating and the remaining part is lost to the environment. The bipropellants enter the motor at a temperature T_i' and gasify at the efficiency of η_{FVR} for fuel, and η_{OVR} for oxidizer, respectively. The overall combustion efficiency is η_{CR} . The subscript R reflects the inclusion of the radiation effects during the phase change and propellant combustion. The heat released by combustion raises the temperature of the gas product, T_C , and partly transfers heat to the wall by conduction q_C and radiation q_R . Thus, T_C is, in general, lower than the adiabatic flame temperature. The flame emission from the hot combustion zone is assumed to be uniformly distributed in the chamber and is transported to the wall, held at the temperature T_W .

2.1.1 Scaling Law of Rocket Motor Performance

The overall energy balance of a rocket motor is given by

$$\begin{aligned} \dot{m}_f(1+\phi) C_P(T_C - T_i') &= -q_C - q_R + \eta_C \dot{m}_f(1+\phi) Q_C \\ &- \dot{m}_f (\eta_{VFR} L_F + \phi \eta_{VOR} L_O) \end{aligned} \quad (1)$$

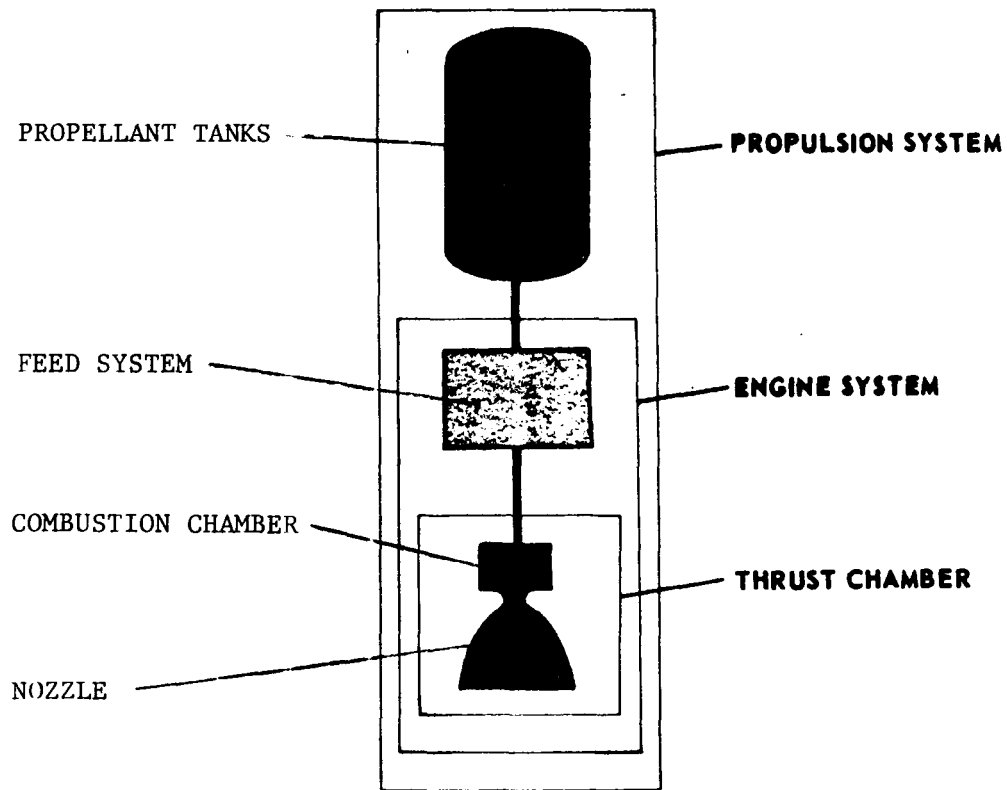


Fig. 1(a). Schematic of Liquid Propulsion System

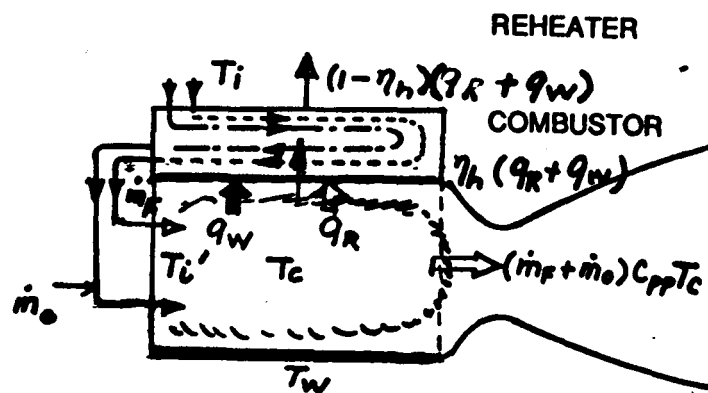


Fig. 1(b). Schematic of Thrust Chamber

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

and for the reheater

$$\eta_h (q_C + q_R) = \dot{m}_f (1+\phi) C_P (T_i' - T_i) \quad (2)$$

where the wall heat transfer rates q_C and q_R are

$$q_C = \frac{\pi N_u k L}{2} (T_C - T_W) \quad (3)$$

$$q_R = \sigma \pi D L (\epsilon_g T_C^4 - a_W T_W^4) \quad (4)$$

ϕ is the mixture ratio, N_u is the Nusselt number, k is the thermal conductivity of the gas, ϵ is the emissivity, σ is the Stefan-Boltzmann constant, a is the surface absorptivity, D is the diameter, and L is the length of the chamber, respectively.

By eliminating T_i' between equations (1) and (2) and by adopting a non-dimensionalization of the resulting equation with appropriate reference properties, one obtains the following algebraic scaling law:

$$\gamma \theta_C^4 + \theta_C = K \quad (5)$$

where

$$\gamma = \frac{\beta (1 - \eta_h)}{1 + \alpha (1 - \eta_h)} \quad (6)$$

$$K = \frac{1 + \theta_i + \alpha (1 - \eta_h) \theta_w + \beta_w (1 - \eta_h) \theta_w^4}{1 + \alpha (1 - \eta_h)} \quad (7)$$

$$\alpha = \frac{N_u k \pi L}{2 C_P \dot{m}_f (1+\phi)} = \frac{N_u k \pi L C^*}{2 C_P P_C A_t} \quad (8)$$

$$\beta = \frac{\sigma \epsilon \pi D L Q_E^3}{C_P^4 \dot{m}_f (1+\phi)^4} = \frac{\sigma \epsilon \pi D L Q_E^3 C^*}{C_P^4 P_C A_t (1+\phi)^3} \quad (9a)$$

$$\beta_W = \frac{\sigma a_W \pi D L Q_E^3}{C_P^4 \dot{m}_f (1+\phi)^4} = \frac{\sigma a_W \pi D L Q_E^3 C^*}{C_P^4 P_C A_t (1+\phi)^3} \quad (9b)$$

$$\theta = \frac{(1+\phi) C_P T}{Q_E} \quad (10)$$

$$Q_E = (1+\phi) \eta_C Q_C - (\eta_{VFR} L_F + \phi \eta_{VOR} L_O) \quad (11)$$

in which C_P is the specific heat of the combustion product, C^* is the characteristic exhaust velocity, A_t is the throat area, L_F and L_O are the latent heat of vaporization of the fuel and oxidizer, respectively, ϕ is the mixture ratio, P_C is the chamber pressure, and k is the thermal conductivity of the gas.

2.1.2 Conductive and Convective Heat Transfer

The conductive and convective heat transfer in rocket motors has been correlated by various semi-empirical expressions involving the Nusselt number. For example, the Colburn equation gives an expression of the Nusselt number in terms of the Reynolds number and Prandtl number. In some sophisticated empirical laws, the effects of the boundary layer temperature gradient on the various gas properties near the chamber wall are accounted for. For example, Bartz (1958) gives the following semi-empirical formula:

$$Nu = \frac{0.026 D^{0.8}}{k} \left(\frac{C_P \mu^{0.2}}{P_r^{0.6}} \right) (\rho v)^{0.8} \left(\frac{\rho_m}{\rho_L} \right) \left(\frac{\mu_m}{\mu_c} \right)^{0.2} \quad (12)$$

The subscript m refers to properties evaluated at the arithmetic mean temperature of the local free-stream static temperature and wall temperature, ρ_L is the local gas density of the free stream, v is the local

gas velocity, subscript c refers to mean chamber temperature, and k is the thermal conductivity of the gas.

In the present analysis, all the local values will be approximated by appropriate mean values.

2.1.3 Radiative Heat Transfer

The radiative emission in a combustion chamber is largely due to molecular rotational and interatomic vibrations of heteropolar gaseous molecules such as water vapor, carbon monoxide, carbon dioxide, hydrocarbons, ammonia, and nitrogen oxides. Symmetrical molecules such as hydrogen, oxygen, and nitrogen show no appreciable emission bands in those wave regions important for chamber radiant heat transfer. In a heterogeneous mixture, such as bipropellant sprays, clouds of finely-spread disperse phases, such as soot particles of hydrocarbon-based propellant combustion, the radiation power could increase by a large factor of 2 to 10 times the equivalent homogeneous gas mixture.

The radiant heat transfer to the wall by the hot gas emission is given by

$$q_R = \sigma A (\epsilon_g T^4 - a_w T_w^4) \quad (13)$$

where ϵ_g is the overall gas emissivity, and a_w is the effective surface absorptivity of the wall. If the wall surface temperature T_w is close to the gas temperature T , then a_w approaches ϵ_g . When T_w and T are not too far apart, the following approximation may be used:

$$\epsilon_g = a_w = \epsilon_{gav} (4 + a + b - c)/4, \quad (14)$$

where $\epsilon_{g,av}$ is the gas emissivity evaluated at the arithmetic mean temperature of T and T_w , and the coefficients a , b , and c are given by

$$a = \frac{\partial (\ln \epsilon_g)}{\partial (\ln P_g L_R)} \quad (15)$$

$$b = \frac{\partial (\ln \epsilon_g)}{\partial (\ln T)} \quad (16)$$

$$c = \frac{\partial (\ln a_w)}{\partial (\ln T)} + \frac{\partial (\ln \epsilon_g)}{\partial (\ln T)} \quad (17)$$

where L_R is the radiation path length of the flame and P_g is the gas pressure. Numerical data of the coefficients a , b , and c for CO_2 and H_2O have been published by Hottel (1959).

In the present analysis, the numerical value of absorptivity, a_w , at the wall that appears in equation (4) can be taken to be different from the gas emissivity value ϵ_g . This difference is incorporated in the definitions of the expression of β which contains ϵ and of β_w which contains a_w . Further assumptions concerning the numerical values of ϵ_g and a_w are discussed in sections 3 and 4.

2.1.4 Performance Profile of the Combustion Process

The method of application of the scaling law, equation (5), for the evaluation of combustor performance is described below.

2.1.4.1 Non-Adiabatic and Adiabatic Chamber Temperature

The non-dimensional chamber temperature θ_C is calculated from the scaling law for the prescribed engine design and operating parameters. By using equations (8) to (18), one can, in general, determine the values of α , β , β_w , and Q_E to estimate the numerical values of K and γ from equations (6) and (7). The dimensionless temperature θ_C is then calculated from equation (7) and plotted as a function of γ and K (Fig. 2). The actual temperature T_C is finally determined from equations (10) and (11).

The non-dimensional isothermal lines, i.e. constant value of θ_C , are a family of straight lines that intersect the K -coordinate at the chamber temperature θ_C (Fig. 2). Note that when $\eta_h = 1.0$, the non-dimensional temperature θ_C assumes the value of $1 + \theta_i$ for any value of α and β . This value is equal to the "adiabatic flame temperature."

An examination of the asymptotic behavior of the scaling law at larger and smaller values of γ reveals a rather interesting feature of the non-adiabatic flame temperature profiles. When γ is much smaller than unity, and K is close to a constant value, i.e., α , β , and $\beta_w \ll 1$, the variation in the chamber temperature with respect

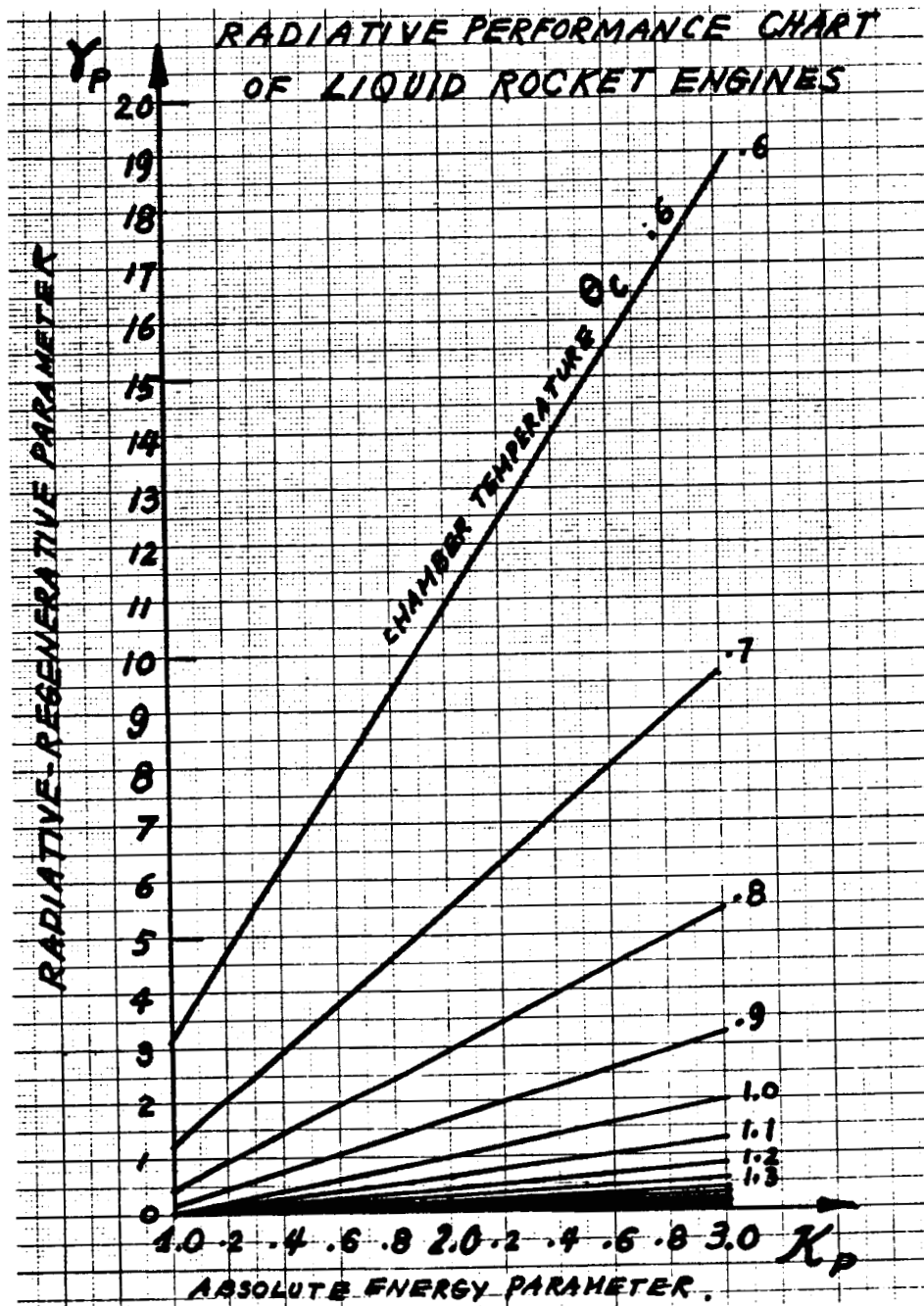


Fig. 2. Scaling Law of Liquid Rocket Combustion Chamber

to the variations in the numerical values of γ , α , β , and β_w is negligibly small. This behavior is similar to what may be termed an "asymptotically adiabatic" pattern wherein the chamber temperature is close to the adiabatic flame temperature in the broad range of the values of α , β , and β_w . Physically, the case corresponds to the rocket engines with a "high enthalpy throughput" in comparison to the combined radiative, conductive, and convective loss, so that the latter thermal energy loss hardly affects the chamber temperature. On the contrary, when the values of γ are not small, the non-dimensional chamber temperature variation, associated with the variations in the values of α , β , and β_w , is not negligibly small, i.e., the chamber temperature depends strongly on the loss mechanisms. This corresponds to those "low enthalpy throughput" engines in which the combined heat loss by radiation, conduction, and convection constitutes a significant fraction of the enthalpy flow rate through the engines.

2.1.5 Engine Classification, High and Low Enthalpy Engines

Two remarkably different chamber temperature characteristics at low γ values ($\gamma \ll 1$) corresponding to high enthalpy flow and high γ ($\gamma \leq 1$) for low enthalpy flow, suggest unique classification of all the liquid rockets into high and low enthalpy engines. In order to illustrate this feature, Table 1 gives a set of data of the SSME and VTE used to estimate the values of γ of each engine.

TABLE 1. VTE and SSME Data

		VTE		SSME	
Combustor	Diameter (ft)	0.204		1.40	
	Length (ft)	0.333		0.67	
Propellants	Propellant	Monomethyl-Hydrazine	Nitrogen Tetroxide	Hydrogen	Oxygen
	State at inlet	Liquid	Liquid	Gas	Liquid
	Boiling Point ($^{\circ}\text{R}$)	649	530	36.9	163
	Heat of vaporization (Btu/lb)	377	178	190	85
	Specific heat (Btu/lb $^{\circ}\text{R}$)	0.7	0.378	3.75	0.5
	Mass flow rate (lb/sec)	0.177	0.2932	150	900
	Inlet temperature ($^{\circ}\text{R}$)	460	460	500	163
	Heat of combustion (Btu/lb)	8910		11250	
Scaling Parameter	γ	$10^{-3} \sim 1^*$		$10^{-5} \sim 10^{-3}^*$	
	K	$1.02 \sim 1.07$		$1.03 \sim 1.04$	

* The upper value is estimated at $N_u = 300$, and the lower value is estimated at $N_u = 3$.

In Table 1, the values of γ for the SSME are much smaller than those of the VTE by an order of magnitude of 10^{-2} to 10^{-3} . This is primarily caused by the significant difference in the total mass flow rate of the SSME, i.e., $\dot{m}_F + \dot{m}_O = 1050$ lb/sec, and the flow rate for the VTE which is 0.470 lb/sec. The latter value amounts to only 0.047 percent of the SSME.

The consequence of the difference in γ values is the dramatic difference in the chamber temperature. As a rule of thumb, the chamber temperature reduction in high enthalpy engines with $\gamma \approx 10^{-3}$, is of the order of "a degree," and that in low enthalpy engines with $\gamma \approx 10^{-1}$ is of the order of $\sim 10^2$ to 10^3 degrees. The detailed chamber temperature loss in the SSME and the VTE is discussed at length in Section 4.

3. PROPELLANT DROPLET VAPORIZATION AND COMBUSTION IN THE RADIATION ENVIRONMENT OF A COMBUSTION CHAMBER

The effects of the radiation enhancement on the propellant vaporization, combustion, overall combustion efficiency, and, subsequently, the chamber temperature, are included in the scaling law using the radiation modulated efficiencies η_{CR} , η_{FVR} , and η_{OVR} . However, the values of these efficiencies can presently only be estimated because of the lack of analytical or experimental data and formulas that interrelate the values of these efficiencies with a combustor design and operating condition.

The basic steps required in the determination of these efficiencies are (1) the acquisition of the droplet laws that account for the radiation enhanced vaporization and combustion, and (2) the prediction of the detailed spray combustion field by a comprehensive CFD code that incorporates the radiation modulated droplet laws. In this section, the first step extends the classical single droplet theory (Williams, 1985) and the second step describes an ad-hoc physical model that interrelates the overall combustion and vaporization efficiencies with the single droplet laws obtained in step 1.

The proposed model of radiation-enhanced droplet vaporization and combustion laws is developed on the basis of the classical single droplet theory with the following additional features. The radiation environment is isotropic and the gas is assumed to be a gray medium. The droplet has a constant absorptivity and the radiative transfer is in radiative non-equilibrium. The state of non-equilibrium radiative heat transfer will therefore require the treatment of the combined radiation-conduction convection mechanism via the solution of radiative transfer equation coupled with the conservation laws of mass, momentum, and energy of the participating gas flow field.

3.1 Mathematical Formulation

The conservation laws governing the droplet processes (Fig. 3) are given by

$$\frac{1}{r^2} \frac{d}{dr} (r^2 \psi_k) = B_k w \quad k = m, o, F, T \quad (18)$$

where

$$\psi_m = \rho v \quad j = F \quad \text{Fuel} \quad (19)$$

ORIGINAL PAGE IS
OF POOR QUALITY

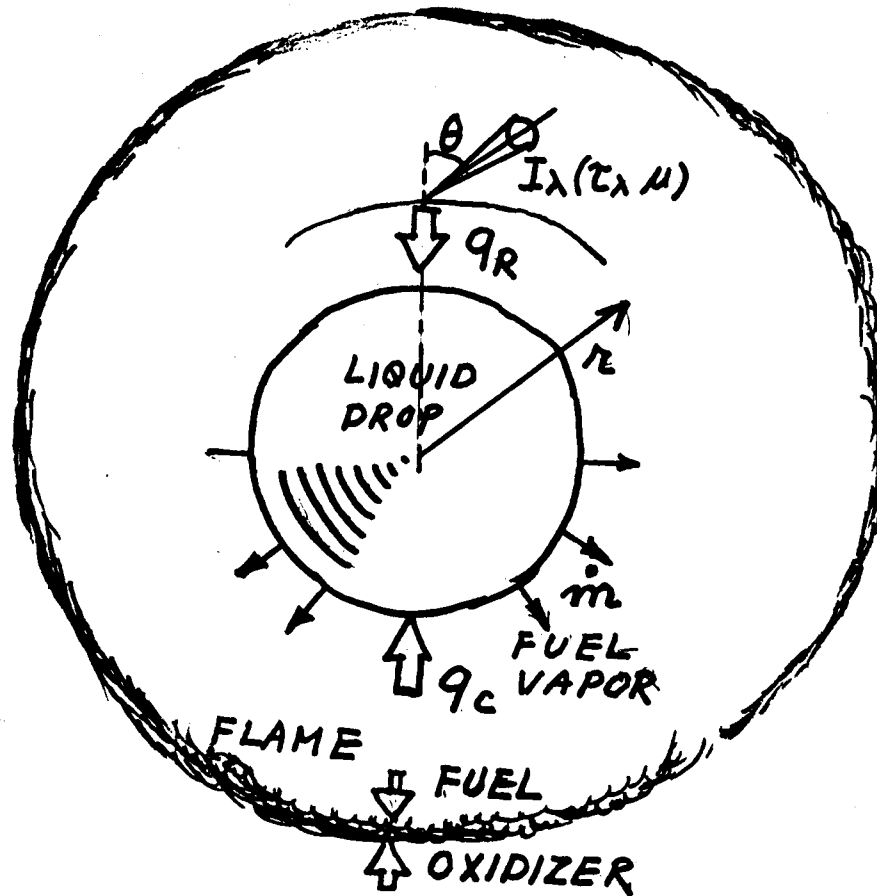


Fig. 3. Single Droplet Combustion in Radiative Environment

$$\psi_j = \psi_m \alpha_j - \rho D \frac{d\alpha_j}{dr} \quad j = 0 \quad \text{oxidizer} \quad (20)$$

$$\psi_T = \psi_m \alpha_T - \frac{\lambda}{c_p} \frac{d\alpha_T}{dr} + q_R \quad (21)$$

in which ρ is the gas density, v is the gas mixture velocity, D and λ are mass diffusivity and thermal conductivity of the gas mixture, respectively, c_p is the gas specific heat at constant pressure, and q_R is the radiative heat flux, α_j and α_T are Schvab-Zeldovich variables defined by

$$\alpha_j = \frac{y_j}{(\nu_j'' - \nu_j') W_j} \quad (22)$$

$$\alpha_T = \frac{\int_{T_h}^T c_p dT}{q^0} \quad (23)$$

$$q_R = \frac{Q_R}{q^0} \quad (24)$$

where y_j is the mass concentration, W_j is the molecular weight of the j th gaseous species, and ν_j'' and ν_j' are the stoichiometric coefficients which appear in the one-step reaction process described as follows:

$$\sum_j \nu_j' m_j = \sum_j \nu_j'' m_j \quad (25)$$

and q^0 is the heat of combustion

$$q^0 = \sum_i h_i^0 W_i (\nu_i' - \nu_i'') \quad (26)$$

in which h_i^0 is the standard heat of formation per unit mass for species i at reference temperature. Q_R and q_R are the dimensional and non-dimensional heat fluxes, respectively.

The radiative heat flux Q_R is given in terms of the radiation intensity I_λ as follows

$$Q_R = 2\pi \int_0^\infty \int_{-1}^1 \mu I_\lambda (\tau_\lambda, \mu) d\mu d\lambda \quad (27)$$

and I_λ is governed by the following radiative transfer equation

$$\mu \frac{\partial I_\lambda}{\partial \tau_\lambda} + \frac{(1-\mu^2)}{\tau_\lambda} \frac{\partial I_\lambda}{\partial \mu} + I_\lambda = (1-w_\lambda) I_{b\lambda} + \frac{w_\lambda}{2} \int_{-1}^1 I_\lambda (\tau_\lambda, \mu) d\mu \quad (28)$$

Here τ_λ is the optical thickness, Ω is the solid angle extended by a surface relative to an observer, λ is the wave length, μ is the $\cos \theta$ with θ being equal to the angle extended by the position vector in radial direction, and that of the radiation beam. The terms a_λ , σ_λ , and β_λ are absorption, scattering, and extinction coefficients, respectively. β_λ and w_λ are given by the following expressions

$$\beta_\lambda = a_\lambda + \sigma_\lambda$$

$$w_\lambda = \sigma_\lambda / \beta_\lambda$$

The subscript b appearing in the radiation intensity $I_{b\lambda}$ refers to that of black body radiation.

The boundary conditions of the present problem are (1) an impermeable droplet surface with respect to oxidizer and combustion product, (2) the balance of the combined radiation and conductive heat transfer with the latent heat of vaporizing species leaving the droplet, (3) intensities of radiation on the droplet surface and at the gaseous environment at $r \rightarrow \infty$, and (4) the temperature and concentration of species at $r \rightarrow \infty$. These conditions are summarized in the following:

$$r = r_\ell :$$

$$\left[\rho v \begin{pmatrix} y_o \\ y_p \end{pmatrix} - \rho D \frac{d}{dr} \begin{pmatrix} y_o \\ y_p \end{pmatrix} \right]_{r_\ell} = 0 \quad (29)$$

$$4\pi r_\ell^2 \left(\frac{dT}{dr} - Q_R \right)_{r_\ell} = \dot{m} L \quad (30)$$

$$I(r_\ell, \mu) = \frac{n^2 \sigma T_b^4}{\pi} \quad (31)$$

$r \rightarrow \infty :$

$$T = T_\infty, \quad Y_i = Y_{i\infty} \quad (32)$$

$$I(\infty, \mu) = \frac{n^2 \sigma T^4}{\pi} \quad (33)$$

3.2 Radiation-Enhanced Vaporization, Burning Laws and Flame Location

By following the mathematical procedure adopted in the classical single droplet theory, one can predict the vaporization or burning rate as the characteristic values of the two-point boundary value problems.

The results of the analysis are listed in the following:

(1) For a vaporizing droplet, the vaporization rate \dot{m}_{VR} is calculated in terms of the "radiation-enhanced vaporization transfer number," B_{VR} , as follows

$$\dot{m}_{VR} = 4\pi \rho D r_\ell \ln(1 + B_{VR}) \quad (34)$$

$$B_{VR} = \frac{\int_{T_L}^T C_p dT}{L} - \frac{\dot{m} q^0}{4\pi (\rho D)^2 L} \int_0^{\xi_\ell} q_R(\xi') \xi'^2 e^{\xi'} d\xi' \quad (35)$$

where

$$\xi = \dot{m} \int_0^r \frac{dr}{4\pi \rho D r^2} \quad (36)$$

(2) For a combustng droplet, the burning rate and the flame location r_f are predicted to be

$$\dot{m}_{CR} = 4\pi \rho D r_\ell \ln (1 + B_{CR}) \quad (37)$$

and

$$\frac{r_f}{r_\ell} = \frac{\ln (1 + B_{CR})}{\ln \left(1 + \frac{W_F \nu_F}{W_o \nu_o} Y_o \right)} \quad (38)$$

where the radiation-enhanced combustion transfer number, B_{CR} , is expressed by

$$B_{CR} = \left[\int_{T_b}^T C_P dT + \frac{q^o y_o}{w_o \nu_o} - \frac{\dot{m} q^o}{4\pi (\rho D)^2} \int_0^{\xi_\ell} q_R (\xi') \xi'^2 e^{\xi'} d\xi \right] / L. \quad (39)$$

Note that when the temperature of the environment is higher than that of the droplet surface, the radiative heat flux points toward the droplet, thus, the q_R has a negative value. Under such circumstances, equations (35) and (39) show that the transfer numbers B_{VR} and B_{CR} are larger than the values determined for the cases when the radiation is absent.

A special case in which $q_R r^2 = \text{constant}$, equations (34) and (35), reduce to the following form:

$$\dot{m}_{VR \text{ or } CR} = 4\pi \rho D r_\ell \ln (1 + B_{VR \text{ or } CR}) \quad (40)$$

where

$$B_{VR} = \frac{B_V}{1 - (Q_R^*/\dot{m}L)} \quad (41)$$

and

$$B_{CR} = \frac{B_C}{1 - (Q_R^*/\dot{m}L)} \quad (42)$$

Here Q_R^* is the rate of the radiative heat transfer to the droplet surface, i.e.,

$$Q_R^* = 4\pi r_\ell^2 Q_R (r = r_\ell) \quad . \quad (43)$$

Thus, the enhancement of the vaporization or combustion of a droplet can be expressed by the transfer number factor given by

$$\epsilon_R = \frac{B_{VR}}{B_V} = \frac{B_{CR}}{B_C} = \frac{1}{1 - (Q_R^*/\dot{m}L)} \quad (44)$$

where ϵ_R is greater than unity.

An alternative expression that describes the vaporization and combustion enhancement is to use the radiation factor η_R defined by

$$\frac{Q_R^*}{\dot{m}L} = \frac{\eta_R \int_{T_b}^T C_P dT}{L} \quad . \quad (45)$$

The vaporization rate can be expressed in terms of the radiation factor η_R as

$$\dot{m}_{VR} = 4\pi \rho D r_\ell \ln \left[1 + \frac{\int_{T_b}^T C_P dT/L}{1 - \eta_R \left(\int_{T_b}^T C_P dT/L \right)} \right] \quad . \quad (46)$$

The droplet vaporization and combustion radiation enhancement factors C_{VR} and C_{CR} can be defined as the ratio of the vaporization and combustion rates as follows:

$$C_{VR} = \frac{\dot{m}_{VR}}{\dot{m}_V} \quad (47)$$

$$C_{CR} = \frac{\dot{m}_{CR}}{\dot{m}_C} \quad (48)$$

where \dot{m}_V and \dot{m}_C are the vaporization and combustion rate for the same environment conditions at $r \rightarrow \infty$, but in the absence of radiative effects on the droplet rate processes.

3.3 Radiation Flux and Intensity

The dimensionless radiation heat flux q_R appearing in the vaporization and combustion rates, equations (34), (35), (37), and (39), is a function of the intensity of radiation I_λ , equation (27). Thus, the determination of the heat flux requires the knowledge of the solution of equation (28).

According to the assumption of a gray gas and an isotropic scattering, the solution of the radiative transfer equation can be expressed in the following power series solution involving Legendre functions

$$I(\tau, \mu) = \sum_{m=0}^{\infty} \frac{2m+1}{4\pi} P_m(\mu) \psi_m(\tau) \quad (49)$$

where $\mu = \cos \theta$. By substituting equation (49) into equation (28), which is specialized to the gray gas approximation, and by equating the coefficient appearing in each order of the Legendre function, one obtains the following set of equations:

$$m = 0 \quad ; \quad \frac{d\psi_1}{d\tau} + 2 \frac{\psi_1}{\tau} + (1-w) \psi_0 = 4\pi (1-w) I_b - \frac{1}{2} w \psi_2 \quad (50)$$

$$m = 1 \quad ; \quad 2 \frac{d\psi_2}{d\tau} + 6 \frac{\psi_2}{\tau} + \frac{d\psi_0}{d\tau} + 3 \psi_1 = 0 \quad (51)$$

$$m = n \quad ; \quad (n+1) \left[\frac{d\psi_{n+1}}{d\tau} + (n+2) \frac{\psi_{n+1}}{\tau} \right] + (2n+1) \psi_n + n \left[\frac{d\psi_{n-1}}{d\tau} - (n-1) \frac{\psi_{n-1}}{\tau} \right] = 0 \quad (52)$$

The net radiation flux q_R^* and incident radiation G can be expressed in terms of the integral as follows:

$$q_R^* = 2\pi \int_{-1}^1 \mu I(\tau, \mu) d\mu \approx \psi_1 \quad (53)$$

$$G^* = 2\pi \int_{-1}^1 I(\tau, \mu) d\mu \approx \psi_0 \quad (54)$$

where the first two terms ψ_0 and ψ_1 appearing in equation (49) are chosen to construct the heat flux q_R^* and incident radiation G^* .

Upon substituting equations (53) and (54), one obtains the following equations governing q_R^* :

$$\frac{d^2 q_R^*}{d\tau^2} + \frac{2}{\tau} \frac{d q_R^*}{d\tau} - \left[3(1-w) + \frac{2}{\tau^2} \right] q_R^* = 4\pi(1-w) \frac{d I_b}{d\tau} \quad (55)$$

and

$$q_R^* = -\frac{1}{3} \frac{d G^*}{d\tau} \quad (56)$$

In general, equation (55) must be solved in conjunction with the conservation equations (18) together with the boundary conditions, equations (31) and (33). Such general solutions can be expressed by a standard procedure, and details of such cases will be reported in the future.

The remaining part of this section will be devoted to a special case in which the absorptivity of the gas is virtually zero. This is the case of practical interest when the gas is optically thin.

By putting w to be equal to unity, one finds that the solution satisfying equation (55) and the boundary conditions, equations (31) and (33), is given by:

$$q_R^* = \frac{4 n^2 \sigma r_\ell}{3 \beta r^2} \left(\frac{T^4 - T_p^4}{\frac{1}{\sqrt{3} \beta r_\ell}} \right) \quad (57)$$

and

$$Q_R = \frac{16 \pi \sigma n^2 r_\ell}{3 \beta} \left(\frac{T^4 - T_b^4}{1 - \frac{1}{\sqrt{3} \beta r_\ell}} \right) \quad (58)$$

The above expression, equation (58), can be substituted into equations (44), (47), and (48) to predict the transfer number enhancement factors, ϵ_R , C_{VR} , and C_{CR} , as well as the radiation factor, equation (45).

4. PERFORMANCE ANALYSIS OF SELECTED LIQUID ROCKET ENGINES

A numerical analysis for the SSME and VTE has been conducted to compare the performance characteristics of high and low enthalpy engines and to assess the need of regeneration of the heat transported to the chamber wall of each engine.

In order to achieve the objectives, the sensitivity of each engine on the following factors is considered: (1) engine size, (2) Nusselt number, (3) emissivity, (4) regenerative efficiency, and (5) power level.

The following two sets of assumptions are adopted as the reference for the performance comparison:

(1) The overall vaporization efficiencies of bipropellants and the overall combustion efficiency, in the absence of radiation, are given by

$$\text{VTE : } \eta_{VF} = 0.94 \quad \eta_{VO} = 0.95 \quad \text{and} \quad \eta_C = 0.92 \quad .$$

$$\text{SSME : } \eta_{VF} = 0.95 \quad \eta_{VO} = 0.9 \quad \text{and} \quad \eta_C = 0.92 \quad .$$

(2) The overall vaporization efficiencies for fuel and oxidizer droplets and the global combustion efficiency in the liquid rocket engine combustion chamber are given by

$$\eta_{VFR} = \frac{\dot{m}_{VFR}}{\dot{m}_{VF}} \eta_{VF} = C_{VFR} \eta_{VF} \quad (59)$$

$$\eta_{VOR} = \frac{\dot{m}_{VOR}}{\dot{m}_{VO}} \eta_{VO} = C_{VOR} \eta_{VO} \quad (60)$$

where C_{VFR} and C_{VOR} are obtained from equations (47) and (48), respectively. These relationships are consistent with the classical criterion of liquid rocket engine design advocated by Priem (1958). For example, C_{VR} 's are determined by

$$C_{VR} = \frac{\ln \left\{ \frac{1 + (1-\eta_R) [C_P(T_m - T_b)/L]}{1 - [\eta_R C_P(T_m - T_b)/L]} \right\}}{\ln \left[1 + \frac{C_P(T_m - T_b)}{L} \right]} \quad (61)$$

The mean gas temperature, appearing in equation (61), is approximated by

$$T_m = \frac{1}{2} (\zeta_c T_c + \zeta_w T_w) \quad (62)$$

in which ζ_c and ζ_w are the weighing factors, respectively, depending on the structure of bipropellant combustion (Chiu, 1986a, 1986b). For example, large group flame emission is largely responsible for the heat transfer to the wall and to the upstream regions to enhance droplet vaporization. Envelope flames of individual droplets and droplet clusters serve to self-accelerate the gasification process. In the latter cases, the numerical value of ζ_c of equation (62) is close to 2 and ζ_w is zero. Vaporizing droplets or clouds near the wall will have ζ_w values close to 2 and ζ_c close to zero. The nature of the distributions of the group flame and drop or cluster bound flames strongly depends on the injection characteristics, droplet mixing pattern, and various spray-turbulent interaction processes. In view of this complex mechanism and the lack of experimental data or numerical simulation at the present time, the average values of ζ_c and ζ_w will be taken to be unity, i.e.,

$$T_m = \frac{1}{2} (T_c + T_w) .$$

4.1 Results and Discussion

The scaling laws and gasification rate models are used to develop a numerical algorithm (see Appendix A), to aid in the systematic numerical prediction of the chamber temperature as a function of chamber design and operating conditions. Basic engine data are listed in Table 1.

4.1.1 Effects of Chamber Geometry

Since the radiative heat loss is approximately linearly proportional to the chamber volume, whereas the conductive heat loss increases linearly with respect to the chamber surface area, the characteristics of chamber temperature loss with respect to the geometry are investigated. Figure 4 shows that the increase in the effective length and diameter of the VTE combustion chamber reduces the chamber temperature appreciably. Note that the maximum loss occurs at the highest gas emissivity, i.e., $\epsilon_g = 1.0$ and $\eta_n = 0$. The maximum loss in a prototype engine is approximately 100°R. However, when the effective length is increased by 10 times the real combustion value, i.e., $L_{eff} = 3.33$ ft, the maximum temperature

ORIGINAL PAGE IS
OF POOR QUALITY

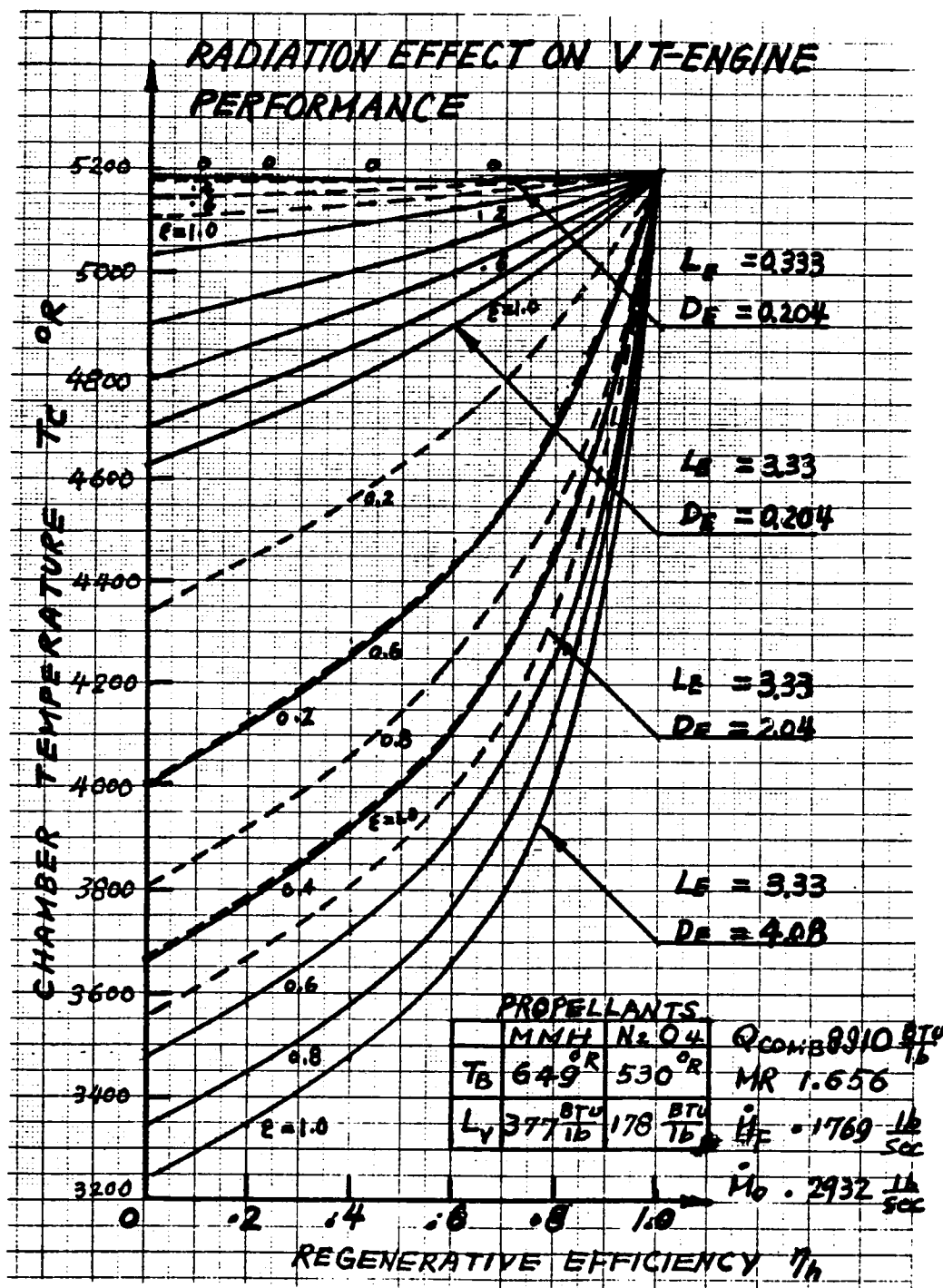


Fig. 4. Radiation Effect on the Performance of the VTE and Additional Hypothetical Sizes

reduction is approximately 600°R. The ratio of the radiative loss to the conductive heat loss is approximately 2 to 1.

In contrast to the VTE, the chamber temperature reduction in SSME is negligibly small. For example, the maximum chamber temperature reduction in a prototype engine, with the emissivity being equal to unity and a vanishing regenerative efficiency, amounts to a few degrees. This is in sharp contrast to VTE with a temperature reduction of 100°R as previously described. Figure 5 shows that the maximum temperature drop of 150°R could occur in a hypothetical SSME engine that has a 200 times larger chamber volume than the prototype.

These significantly different chamber temperature loss characteristics of the VTE and SSME are attributed to the low and high enthalpy flow characteristics of each engine.

Figure 6 shows the ranges of the variations of γ and K for two types of engines, the VTE and the SSME. Note that the variations in the values of γ for the SSME range from 10^{-5} to 10^{-3} , and K remains nearly constant as the engine size increases. Note that the values of γ in the VTE are approximately 100 times greater than that of the SSME and a pronounced variation in K value occurs when the engine size increases. The latter factor has a small positive effect for minimizing the chamber temperature reduction.

The results shown above and in the remaining part of this section suggest that the low-enthalpy engines could have a relatively large temperature drop due to radiative heat loss. The regenerative scheme will be effective in particular when such engines are designed or being operated in marginal conditions such as lower power level flight.

4.1.2 Effects of Nusselt Number

The relative importance of the conductive-convective heat loss to that of radiation is assessed by comparing the chamber temperature variations in a range of Nusselt number extending from an order of unity to 10^3 . Figure 7 shows that the variation in T_c for the VTE is relatively small in the range of $1 < N_u < 10^2$. A significant conductive-convective loss occurs when N_u exceeds a critical value of approximately 300. The presence of such a critical Nusselt number suggests that the low-enthalpy engines, such as the VTE, may be further classified into the "radiation loss dominated engine" when N_u is smaller than the critical value and "conduction-convection loss dominated engine" if N_u is greater than the critical value.

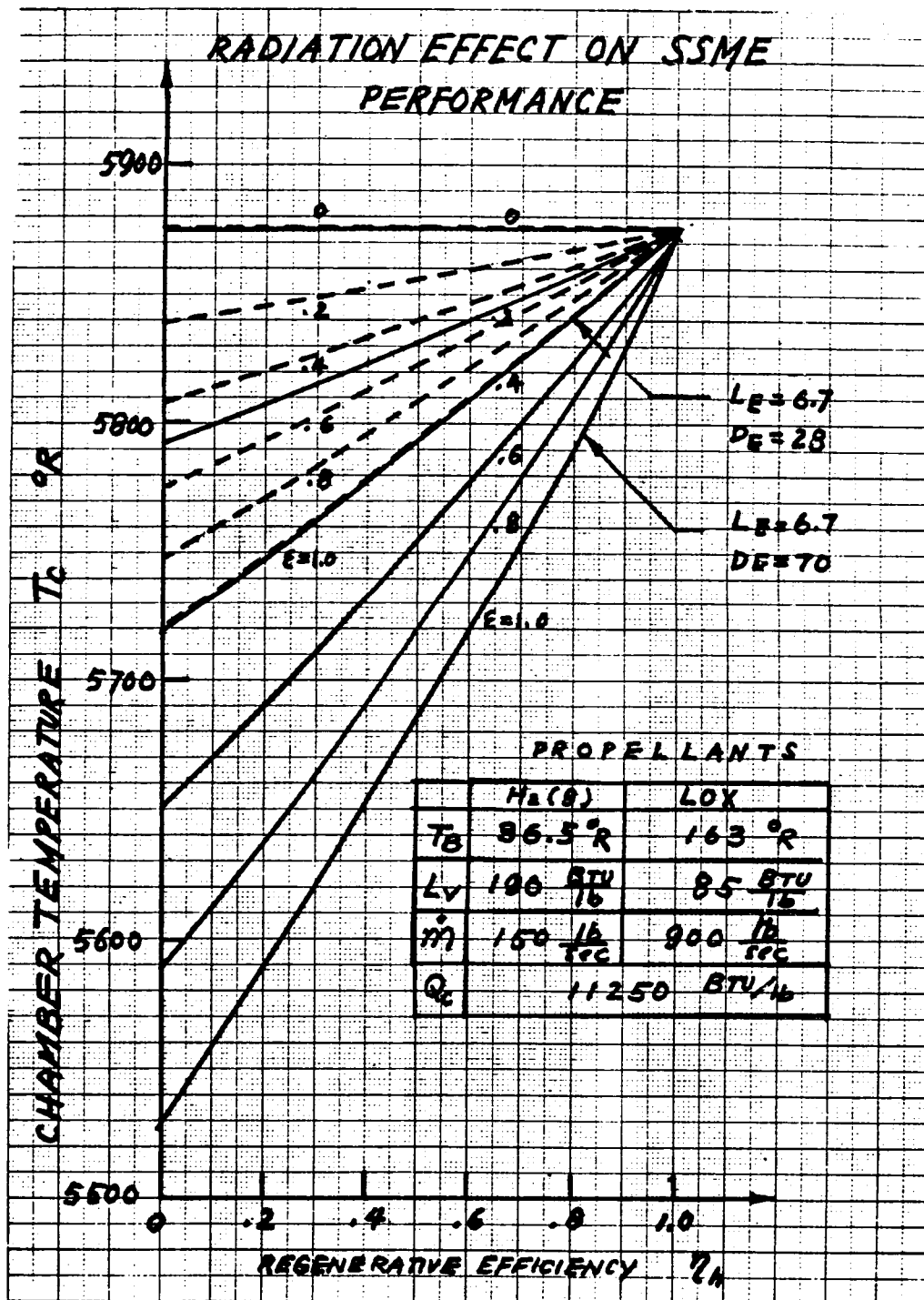


Fig. 5. Radiation Effects on the Performance of the SSME and Additional Hypothetical Sizes

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

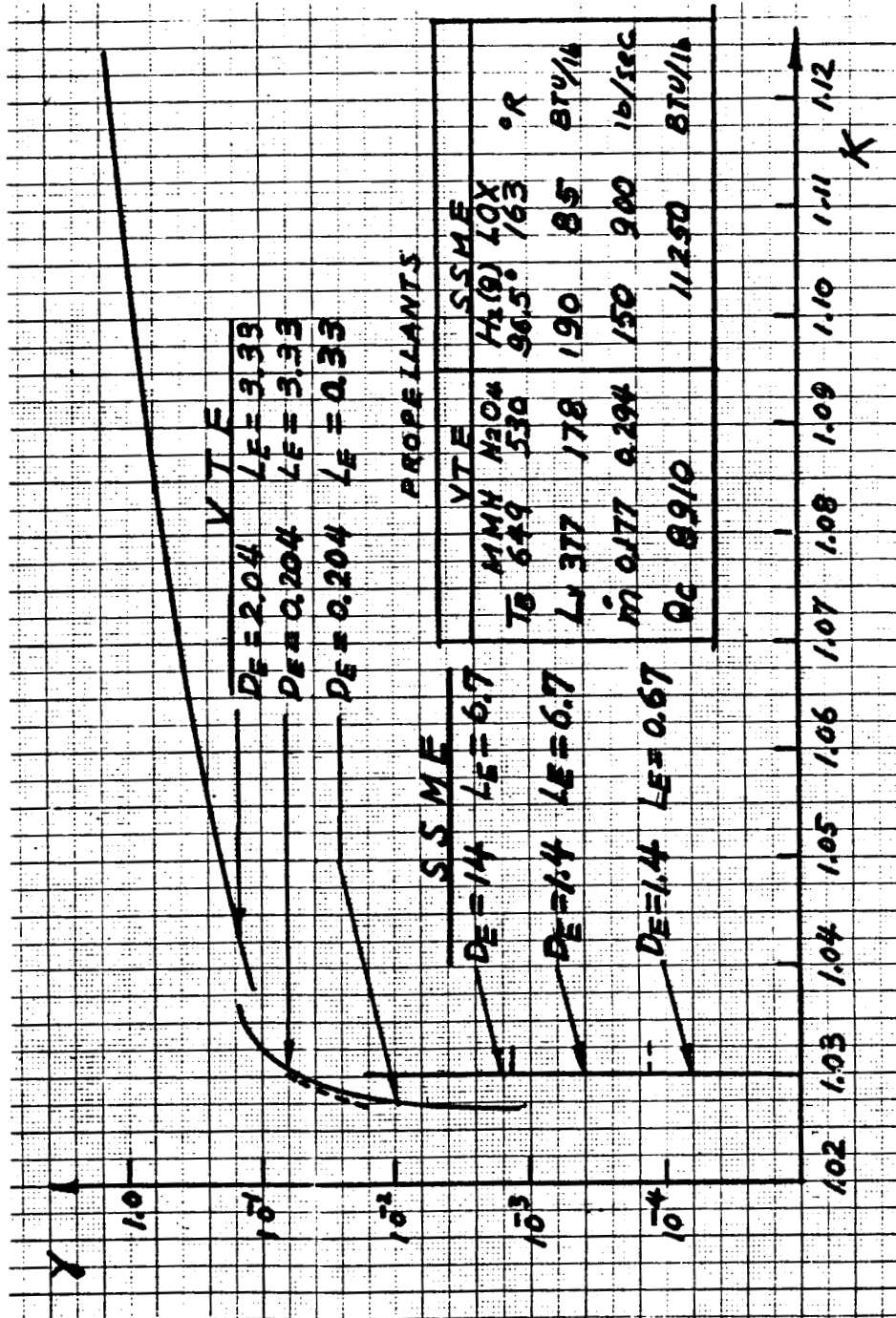


Fig. 6. Scaling Laws of the VTE and SSME of Various Sizes

ORIGINAL PAGE IS
OF POOR QUALITY

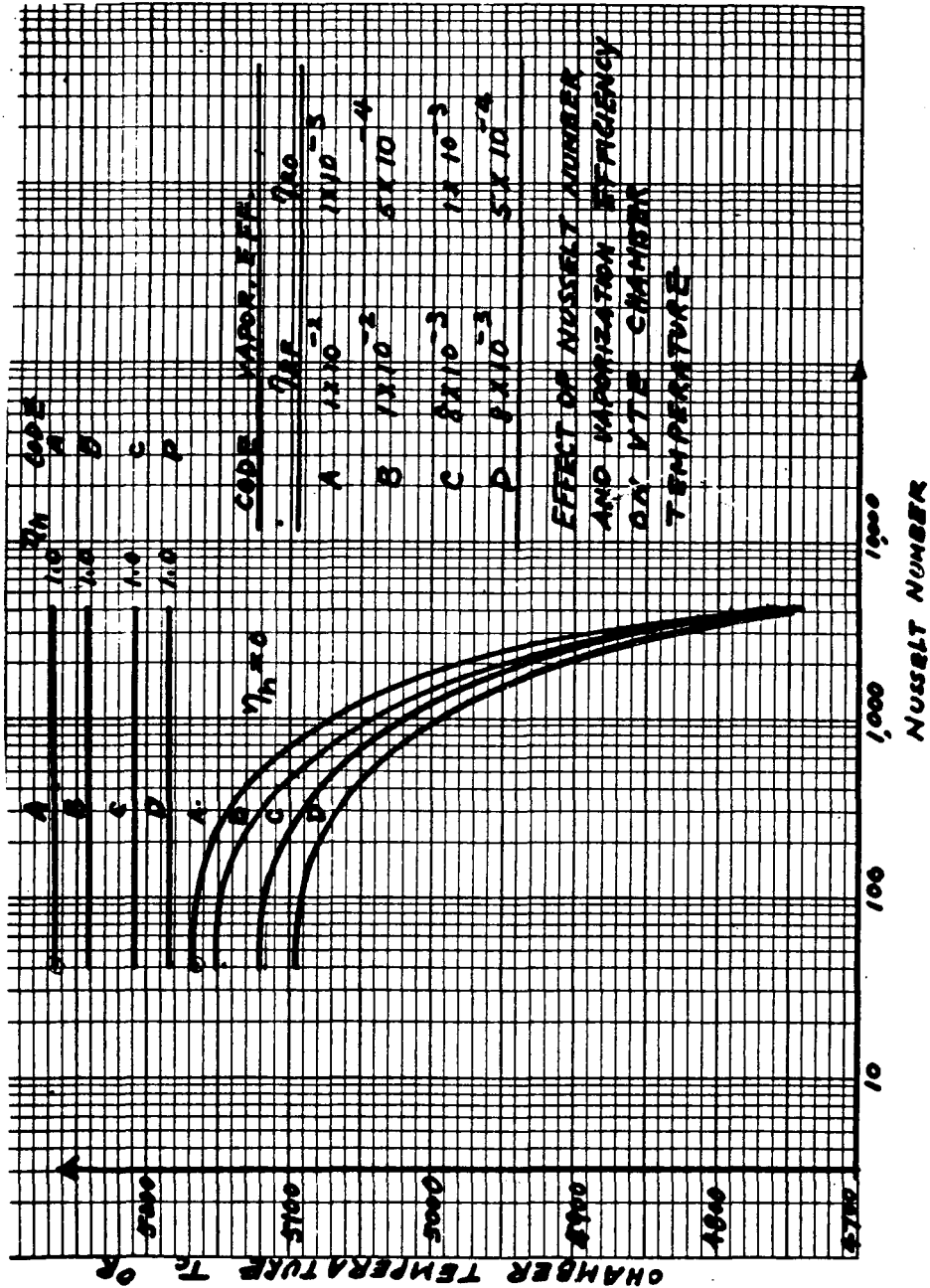


Fig. 7. Effect of Nusselt Number and Vaporization Efficiency on VTE Chamber Temperature

In contrast to the VTE, the chamber temperature losses estimated for the SSME in the same range of Nusselt numbers are found to be negligibly small (Fig. 8).

4.1.3 Effects of Radiation-Enhanced Combustion and Vaporization Rates

The effects of radiation of the propellant gasification enhancement in the SSME is shown in Figure 9. Note that an almost 150°R gain in the chamber temperature is achieved in an adiabatic flame temperature case when η_{RO} is increased from 5×10^{-4} to 2×10^{-3} . The gain in chamber temperature at various degrees of regeneration is approximately 150°R for the prescribed increase in η_{RO} . Figures 7 and 8 also show the general feature of the chamber temperature gain in the range of the Nusselt numbers considered.

This general temperature gain characteristic is attributed to the fact that the vaporization enhancement rate is approximately proportional to the enthalpy flows and, thus, constitutes a significant fraction of energy addition into the combustion process.

Thus, for both high- and low-enthalpy engines, the strategy is to use as much radiation energy as possible to enhance the gasification process and the associated combustion efficiency while minimizing the radiative transfer loss to the wall.

4.1.4 Effects of Engine Power Level

Since the throttling of the engine power output is accomplished, in general, by the reduction in total enthalpy flow, the lower power level operation will exhibit low enthalpy characteristics; i.e., high γ engine behavior. Figure 10 shows the VTE chamber temperature variations in the ranges of power level and gas emissivity between 0 to 1.0. The effects of the Nusselt number, the power level and gas emissivity on the chamber temperature are shown in Figure 11.

The results once again suggest that a dramatic reduction in chamber temperature, on the order of a few to several hundred degrees, is expected in low enthalpy engines.

ORIGINAL PAGE IS
OF POOR QUALITY

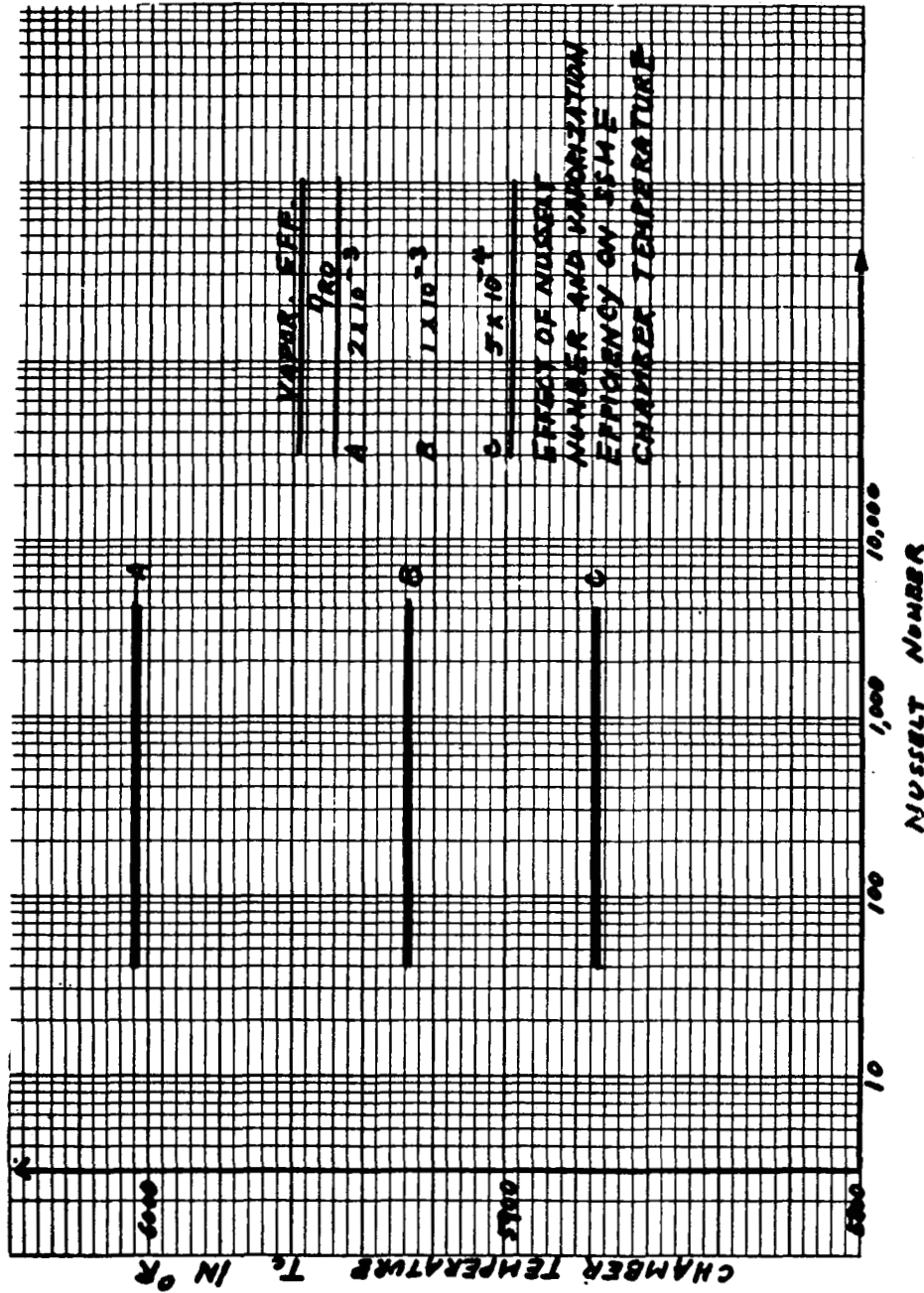


Fig. 8. Effect of Nusselt Number and Vaporization Efficiency on
SSME Chamber Temperature

ORIGINAL PAGE IS
OF POOR QUALITY

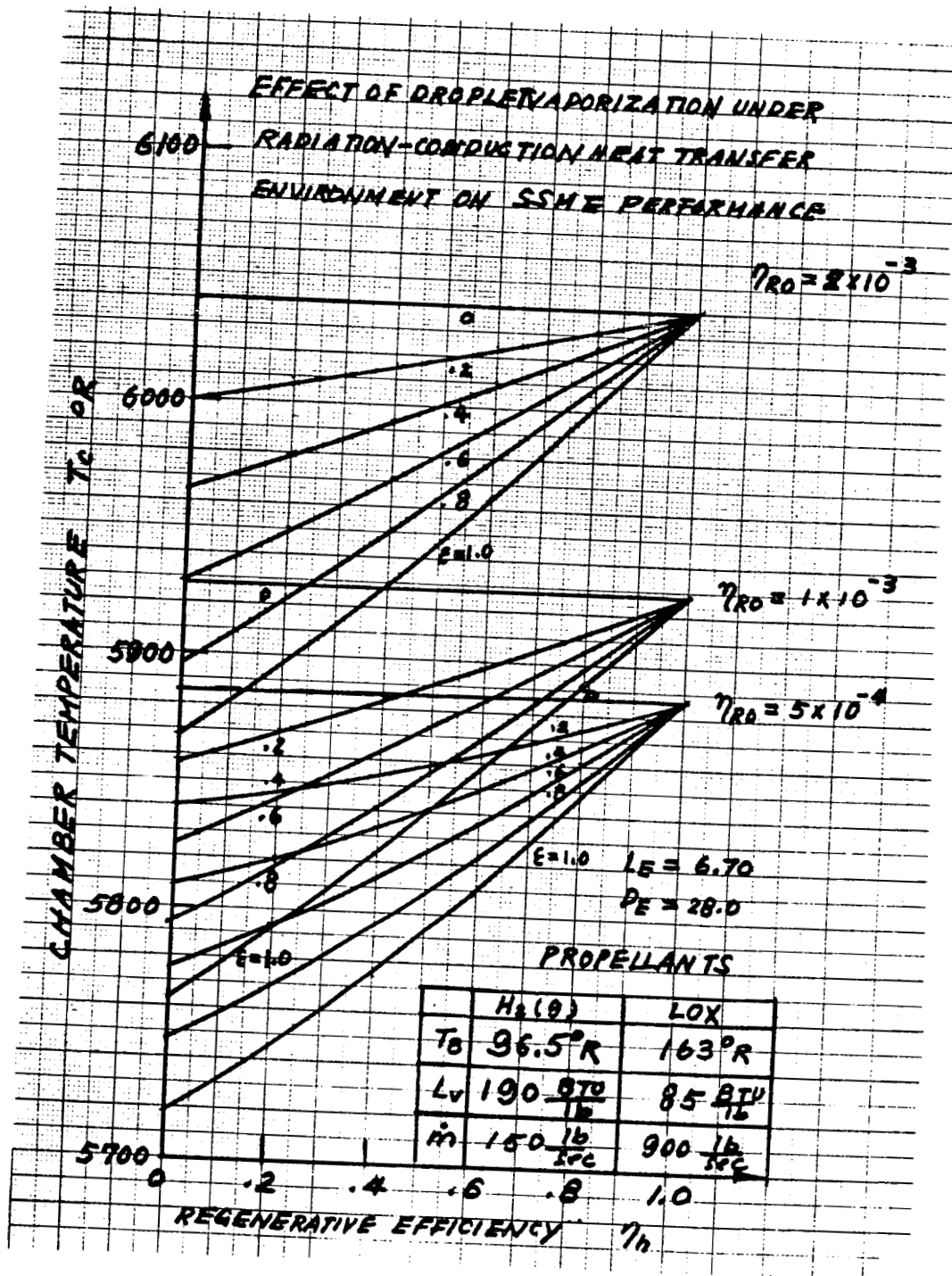


Fig. 9. Effect of Droplet Vaporization Under Radiative-Conductive Heat Transfer Environment on SSME Performance

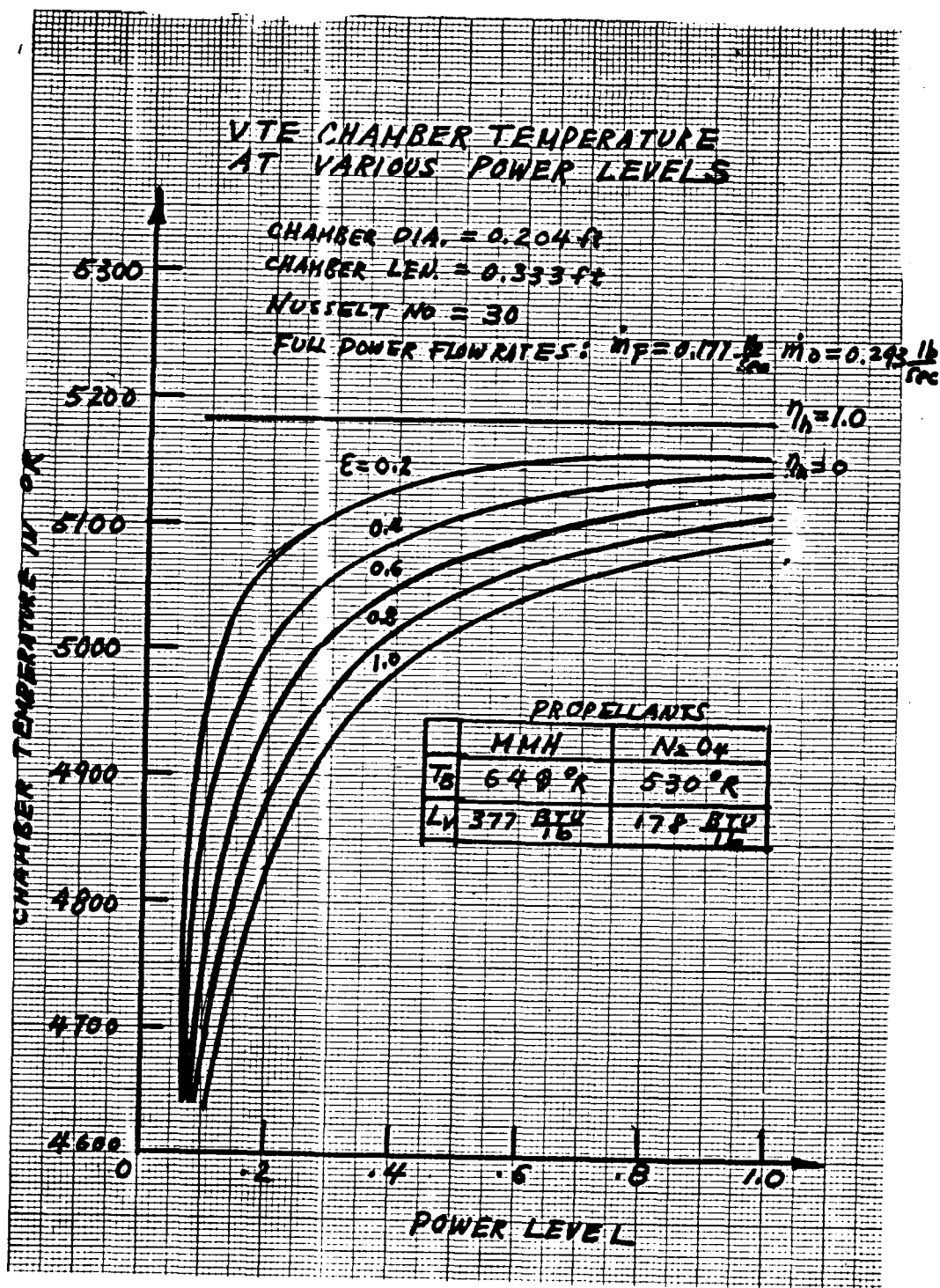


Fig. 10. Effect of Radiation on VTE Chamber Temperature
at Various Power Levels

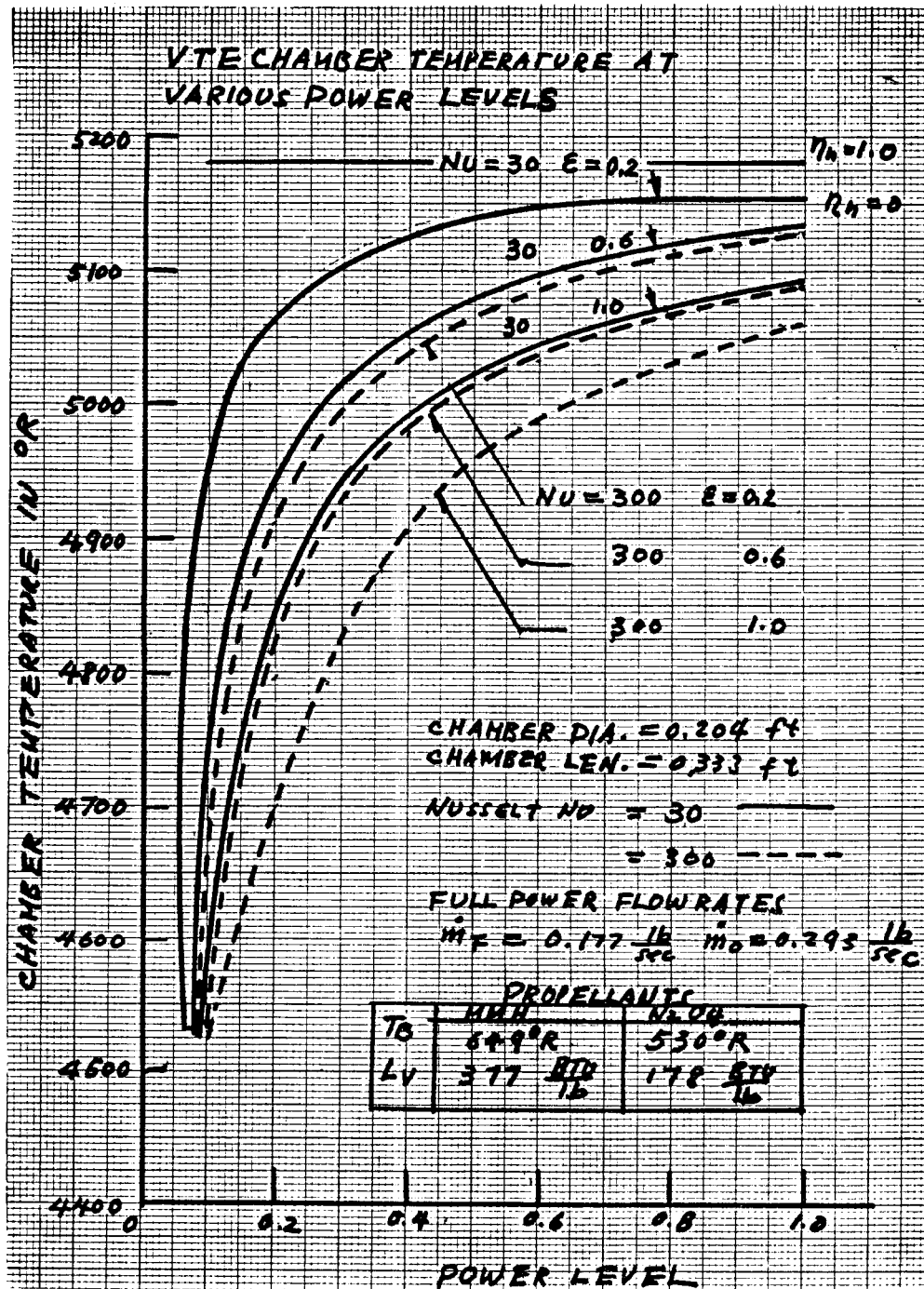


Fig. 11. Effect of Radiation on VTE Chamber Temperature at Selected Nusselt Numbers and at Various Power Levels

5. CONCLUSION

The scaling law coupled with the radiation-enhanced propellant combustion model and associated computational algorithms have been developed to assess the impacts of radiation on the chamber temperature of liquid rocket engines. Two major impacts due to radiation are (1) the reduction in the temperature due to unrecovered radiative conductive and convective heat transfer to the chamber wall with a radiative amount of approximately 70% of the total loss when the Nusselt number is smaller than a few hundred, and with increasing dominance by a conductive-convective loss at larger Nusselt numbers, and (2) the gain in the chamber temperature due to the radiation-enhanced gasification rate that tends to increase the combustion efficiency by a few percent over the non-radiating combustion environment.

Based on the characteristics of radiative loss and gain, as presented in the numerical illustration, liquid rocket engines can be classified into high-enthalpy engines, such as the SSME; and low-enthalpy engines, such as the VTE. The performance of the latter type of engines can be significantly impaired by radiative loss at all power levels with high emissivity due to heteropolar gaseous molecules such as H_2O , CO_2 , CO , NH_3 , and NO_2 . A regenerative system with an efficiency higher than 50% would be required to maintain an acceptable engine performance.

Although the overall combustion efficiency model presented here is crude, and uncertainties of the numerical values adopted may affect the accuracy, the positive impact of radiation-enhanced droplet vaporization appears to be a fundamental consideration in the selection of the propellant and the design of the injector. Much experimental data and extensive modeling, especially with CFD, are required to support the preliminary discoveries.

The overall assessment of the general problem of radiation phenomena in liquid rocket engines clearly indicates the need for research in the following specific problem areas.

(1) Droplet-radiation Interaction — The problems related with the radiative effects on the droplet ignition, combustion, extinction, as well as droplet saturation, condensation of cryogenic and non-cryogenic propellants, are of basic importance in radiation-coupled spray combustion. The effects of the collective interaction phenomena can significantly alter the radiation beam penetration, scattering, absorption, and emission. Droplets near walls, in envelope flames, and recirculation zones are subject to specific radiation view factors which can promote or impair the gasification rates.

(2) Turbulence-Radiation Interaction — High temperature, high intensity turbulence in the combustion zone provokes turbulence-radiation interaction when the radiative cooling takes part in viscous dissipation and heating. In particular, when the wavelength of the radiation is comparable to the scale of the turbulence, the intercoupling could contribute strongly to the radiative decay of turbulence. Many of these effects are anticipated to depend on the molecular properties, the chamber pressure, and temperature. Heterogeneous mixtures, soot particles and droplets which have high emissivity and absorptivity are also important factors in the local turbulence-radiation interaction.

(3) Flame-Radiation Interaction — Thermal emission and chemiluminescence in a flame zone are the cause of the radiative cooling in optically thin flames. The cooling impairs the flame stability characteristics. Many of these phenomena depend on the molecular and thermochemical properties of the gas mixture in the preheating and combustion zones. In assessing the interaction between the flame and radiation, one needs to differentiate premixed versus diffusion/mixing controlled flames because the heat release rate and the cooling rate in each type of flame are characteristically different. This suggests that the flame extinction characteristics in the two types of flames are different. Additionally, the presence of soot particles in the flame environment affects the emissivity and absorptivity of the mixture. Systematic studies of these fundamental processes under typical rocket engine operating conditions are still lacking at present.

The prediction of combustion processes in a radiative environment requires an advanced spray combustion CFD code that is coupled with a radiative transfer subcode. The latter accounts for the non-equilibrium radiation transport processes. In addition, combustion submodels including radiative droplet laws, atomization processes, turbulence models, gas phase combustion, as well as emissivity, absorptivity models in super or subcritical states under radiation effects must be supplemented. Finally, new computational techniques based on finite difference schemes for conventional CFD and special computational techniques such as multi-flux schemes must be effectively incorporated. In addition, the initial and boundary conditions must be supplied to formulate a well-posed computational problem.

REFERENCES

- Afgan, N. H. and J. M. Beer, Editors, Heat Transfer in Flames, John Wiley & Sons, New York, NY (1974).
- Bartz, D. R., "Factors Which Influence the Suitability of Liquid Propellants as Rocket Motor Regenerative Coolants," Jet Propulsion, Vol. 28, pp. 46-53 (Jan. 1958).
- Chiu, H. H., "Theory of Bipropellant Combustion, Part I," AIAA-86-0220, AIAA 24th Aerospace Sciences Meeting, Reno, NV (1986a).
- Chiu, H. H., et al., "Analytical Prediction of Combustion Performance Characteristics of Bipropellant Liquid Rocket Engine Combustor," JANNAF Combustion Meeting (1986b).
- Hottel, H. C., "Engineering Calculations of Radiant Heat Exchange," Turbulent Flows and Heat Transfer, C. C. Lin, Editor, Vol. V, High Speed Aerodynamics and Jet Propulsion, Princeton University Press, Princeton, NJ (1959).
- Priem, R. J., "Propellant Vaporization as a Criterion for Rocket Engine Design," NASA TN 3883 (1958).
- Sutton, G. P. and D. M. Ross, Rocket Propulsion Elements, John Wiley & Sons, New York, NY (1975).
- Summerfield, M., "The Liquid Propellant Rocket Engine Liquid Propellant Rockets," Editor D. Altman, No. 1., Princeton Aeronautical Paperbacks, Princeton University Press, Princeton, NY (1960).
- Williams, F. A., Combustion Theory, Benjamin/Cumming Publishing Co., Inc., Menlo Park, CA (1985).

APPENDIX A
PROGRAM LISTING

ORIGINAL PAGE IS
OF POOR QUALITY

```

C   PAUL CHIU'S RADIATION COMPUTER MODEL
      DIMENSION ETAREG(11),GRAD(10),INCL(2),ETARF(10),ETARO(10)      ANK 8/88
      DATA FTLBTU/778.16/,GC/32.174/,KG,LE,MF,NO/4*1/,PI/3.141592653/, ANK 8/88
1    IP,IRAD/0,0/,MAX,MAXR/2*50/,TOL/0.0020/,INCL/'EXCL','INCL'/      ANK 8/88
      REAL ITLIM,KCOND,KP,LENGTH,LF,LO,MDOTF,MDOTO,NUSSEL            ANK 7/88

C   NAMELIST /PAUL/ CPFI,CPOI,CPPROD,DIAM,ETAC,ETAREG,ETARF,ETARO,    /PAUL/
1    ETAVF,ETAVO,ETAWAL,GRAD,IP,IRAD,ITLIM,KCOND,KG,LE, /PAUL/
2    LENGTH,LF,LO,MAX,MAXR,MDOTF,MDOTO,MF,NO,NUSSEL, /PAUL/
3    QCOMB,SIGMA,TBF,TBO,TFI,TOI,TOL,TWALL,ALPHA,PHI,QE /PAUL/

C   READ (5,PAUL)                                                    ANK 7/88
      PHI = MDOTO/MDOTF                                              ANK 7/88
      PHI1 = 1.0 + PHI                                              ANK 7/88
      QE = ETAC*PHI1*QCOMB - ETAVF*LF - ETAVO*LO*PHI                ANK 7/88
      ALPHA = KCOND*LENGTH*NUSSEL*PI/(2.0*CPPROD*FTLBTU*GC*MDOTF*PHI1) ANK 7/88
      WRITE (6,2) INCL(IRAD+1)                                       ANK 8/88
2    FORMAT (//20X,'PROPELLANT VAPORIZATION AFFECTED BY REGENERATED HEAANK 8/88
1T (ETAREG) ',A4,'UDING RADIATION'//)                                ANK 8/88
      WRITE (9,2) INCL(IRAD+1)                                       ANK 8/88
      WRITE (6,PAUL)                                                 ANK 7/88
      WRITE (9,PAUL)                                                 ANK 8/88
      DO 40 M = 1,MF                                                 ANK 8/88
      DO 40 N = 1,NO                                                 ANK 8/88
      DO 40 K = 1,KG                                                 ANK 8/88
      WRITE (9,6)                                                    ANK 8/88
6    FORMAT (/)                                                      ANK 8/88
      DO 40 L = 1,LE                                                 ANK 8/88
      J = 0                                                         ANK 8/88
      TC1 = 0.0                                                     ANK 8/88
5    BETA = DIAM*GRAD(K)*LENGTH*PI*QE**3*SIGMA/(3600.0*CPPROD**4*MDOTF ANK 8/88
1    *PHI1**4)                                                      ANK 7/88
      ALPHAP = ALPHA*(1.0 - ETAREG(L))                              ANK 8/88
      BETAP = BETA*(1.0 - ETAREG(L))                                ANK 8/88
      GAMMAP = BETAP/(ALPHAP + 1.0)                                  ANK 8/88
      CPROPI = (CPFI + CPOI*PHI)/PHI1                               ANK 7/88
      TPROPI = (CPFI*TFI + CPOI*PHI*TOI)/(CPFI + CPOI*PHI)         ANK 7/88
      THETAI = CPROPI*PHI1*TPROPI/QE                                ANK 7/88
      THETAW = CPROD*PHI1*TWALL/QE                                  ANK 7/88
      KP = (1. + ALPHAP*THETAW + BETAP*THETAW**4 + THETAI)/(ALPHAP+1.) ANK 8/88
      QC = KP                                                         ANK 7/88
      I = 1                                                         ANK 7/88
10   REST = GAMMAP*QC**4 + QC - KP                                   ANK 7/88
      IF (ABS(REST) .LT. ITLIM .OR. I .GT. MAX) GO TO 30            ANK 7/88
      IF (I .GE. 2) GO TO 20                                         ANK 7/88
      QC1 = QC                                                       ANK 7/88
      REST1 = REST                                                  ANK 7/88
      I = 2                                                         ANK 7/88
      QC = 1.20*QC                                                  ANK 7/88
      GO TO 10                                                       ANK 7/88
20   QC2 = QC1                                                       ANK 7/88
      REST2 = REST1                                                  ANK 7/88
      QC1 = QC                                                       ANK 7/88
      REST1 = REST                                                  ANK 7/88
      IF (ABS(REST2 - REST1) .LE. 0.0) GO TO 30                     ANK 7/88
      QC = QC1 - REST1*(QC2 - QC1)/(REST2 - REST1)                 ANK 7/88
      I = I + 1                                                      ANK 7/88
      GO TO 10                                                       ANK 7/88
30   TC = QC*QE/(CPPROD*PHI1)                                       ANK 7/88
      IF (IRAD .EQ. 0) GO TO 35                                     ANK 8/88
C   RADIATION IS TO BE INCLUDED
      TC2 = TC1                                                      ANK 8/88
      TC1 = TC                                                       ANK 8/88

```

PROGRAM LISTING (Continued)

```

J = J + 1                                ANK 8/88
TRATIO = ABS(TC2 - TC1)/TC1              ANK 8/88
IF (TRATIO .LE. TOL .OR. J .GT. MAXR) GO TO 35 ANK 8/88
TCR = (TC + ETAWAL*TWALL)/2.0            ANK 8/88
AF = CPPROD*(TCR - TBF)/LF               ANK 8/88
AO = CPPROD*(TCR - TBO)/LO               ANK 8/88
CRADF = LOG((1.0+(1.0-ETARF(M))*AF)/(1.0-ETARF(M)*AF))/LOG(1.0+AF) ANK 8/88
C CRADF = 1.0                            FOR SSME
CRADO = LOG((1.0+(1.0-ETARO(N))*AO)/(1.0-ETARO(N)*AO))/LOG(1.0+AO) ANK 8/88
ETACR = (CFADF*CRADO)*ETAC              ANK 8/88
ETAVFR = CFADF*ETAVF                   ANK 8/88
ETAVOR = CRADO*ETAVO                   ANK 8/88
QE = ETACR*PHI1*QCOMB - ETAVFR*LF - ETAVOR*LO*PHI ANK 8/88
GO TO 5                                ANK 8/88
35 IF (IP .GT. 0) GO TO 40              ANK 8/88
WRITE (6,1) I,K,L,CPROPI,ETAREG(L),GRAD(K),KP,QC,TC,TPROPI, ANK 8/88
1 ALPHAP,BETAP,GAMMAP,REST,THETAI,THETAW ANK 8/88
1 FORMAT (/ ' I =',I4, ' K =',I3, ' L =',I3, ' CPROPI =',F9.6, ANK 7/88
1 ' ETAREG =',F5.3, ' GRAD =',F5.3, ' KP =',F9.6, ' QC =',F9.6, ANK 8/88
2 ' TC =',F7.1, ' TPROPI =',F7.1/' ALPHAP=',1PE13.6, ' BETAP=', ANK 8/88
3 E13.6, ' GAMMAP=',E13.6, ' REST=',E13.6, ' THETAI=',E13.6, ANK 7/88
4 ' THETAW=',E13.6/) ANK 7/88
WRITE (6,3) J,M,N,CRADF,CRADO,ETACR,ETAVFR,ETAVOR,QE,TCR,ETARF(M), ANK 8/88
1 ETARO(N),AF,AO,TRATIO ANK 8/88
3 FORMAT ( ' J=',I3, ' M=',I3, ' N=',I3, ' CRADF=',F8.5, ' CRADO=', ANK 8/88
1 F8.5, ' ETACR=',F8.5, ' ETAVFR=',F8.5, ' ETAVOR=',F8.5, ANK 8/88
2 ' QE=',F10.3, ' TCR=',F8.2/' ETARF(M) =',F9.6, ' ETARO(N) =', ANK 8/88
3 F9.6, ' AF =',F10.3, ' AO =',F10.3, ' TRATIO =',E13.6/) ANK 8/88
40 WRITE(9,4) L,ETAREG(L),TC,GRAD(K),ETARF(M),ETARO(N),TCR,QC,QE ANK 8/88
4 FORMAT(/ ' L =',I3, ' ETAREG =',F6.4, ' TC =',F8.2, ' GRAD =',F6.4, ANK 8/88
1 ' ETARF =',F8.6, ' ETARO =',F8.6, ' TCR =',F8.2, ' QC =',F9.6, ANK 8/88
2 ' QE =',F10.3) ANK 8/88
END ANK 7/88

```

INPUT DATA

```

$PAUL
CPFI = 0.70, CPOI = 0.3780, CPPROD = 1.650, DIAM = 0.20412, ETAC = 0.920,
ETAREG = 1.,.9.,.8.,.7.,.6.,.5.,.4.,.3.,.2.,.1.,.0, ETAVF = .94, ETAVO = .95, IRAD = 1,
GRAD = 1.0,0.8,0.6,0.4,0.2,0.0, ITLIM = 1.0E-10, MF = 1, NO = 1,
KCOND = 2.0, KG = 6, LE = 11, LENGTH = 0.33333, LF = 377.0, LO = 178.20,
MDOTF = 0.1415728, MDOTO = 0.2345336, NUSSEL = 300.0, QCOMB = 8910.0,
SIGMA = 1.7120E-9, TFI = 460.0, TOI = 460.0, TWALL = 2500.0, TOL = 0.0020,
ETAWAL = 1.0, ETARF = .008,.01,, ETARO = .0005,.001, TBF = 649.0,
TBO = 530.0,
$END

```


1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

A PERFORMANCE/LOSS EVALUATION OF SSME HPFTP TURBINE

Prepared by

Mingking K. Chyu

Academic Rank:

Assistant Professor

University and Department

Carnegie Mellon University
Department of Mechanical
Engineering

NASA/MSFC:

Laboratory:
Division:
Branch:

Propulsion
Component Development
Turbomachinery and
Combustion Devices, EP62

MSFC Colleague:

George M. Young, III

Date:

August 1988

Contract No.:

NGT 01-002-099
The University of Alabama

A PERFORMANCE/LOSS EVALUATION OF SSME HPFTP TURBINE

M.K. Chyu
Department of Mechanical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

George M. Young III
Propulsion Laboratory
NASA Marshall Space Flight Center
Huntsville, AL 35812

ABSTRACT

An evaluation of component losses of the High Pressure Fuel Turbopump (HPFTP) in the Space Shuttle Main Engine (SSME) is performed using a mean-line prediction method. This is accompanied with an extensive review of loss correlations in the literature. The present prediction uses an existing gas path velocity triangle, real LH_2 and H_2O gas properties and loss correlations selected from the literature. The significant losses incurred in the HPFTP turbine include profile loss, secondary loss and tip clearance loss. Results obtained from the present prediction are compared to those calculated from a quasi-three-dimensional numerical analysis. Except for the clearance loss, the present loss coefficients are in general higher than their counterparts from the quasi-three-dimensional analysis. The fact that the mean-line velocity data being unable to represent actual flowfields near the hub and the tip regions is largely responsible for the uncertainty involved in the present method. On the other hand, due mainly to the ad-hoc nature of the studies involved, the correlations currently available in the literature may not be suitable for accurate loss prediction of the particular rocket turbine in the SSME HPFTP. Further studies particularly in the areas of tip clearance loss, coolant loss, secondary loss and their interactions are desirable. Fundamental phenomena concerning flow unsteadiness in wake shedding and turbulence are also important.

**ORIGINAL PAGE IS
OF POOR QUALITY**

ACKNOWLEDGEMENTS

The first author (MKC) of this report is grateful to NASA-Marshall Space Flight Center and ASEE for the appointment of Summer Faculty Research Fellowship. In addition, he expresses special appreciation to Mr. George M. Young III for his collaboration in this project.

Thanks to Ms. Mary Beth Boyd for her assistance in providing important information and cooperation, Ms. Joan G. Trolinger for her help in collecting references, Dr. Helen McConnaughey for her support in the numerical aspect of this study, and Mr. Loren A. Gross for his enthusiastic interest in this project.

For Professor Mike Freeman and Ms. Ernestine Cothran, their excellent management of the entire program is highly appreciated.

NOMENCLATURE

c	chord length
C_b	base pressure coefficient, see Eq. (7)
C_L	lift coefficient
h	mean blade height
k	tip clearance, see Eqs. (12)-(15)
L	blade surface length in axial direction
M	Mach number
p	pressure
r	ratio of specific heat of constant pressure to that of constant volume, see Eq. (3)
R_T	blade tip radius
R_m	blade mean radius
s	blade pitch
t_e	blade trailing edge thickness
U	circumferential velocity
V	mean velocity in axial direction
Y	pressure coefficient, see Eq. (1)
Z	Ainley-Mathieson loading parameter, see Eq. (10)

Greek Symbols

α	relative flow angle at blade row, see Fig. 1
β	blade angle, see Fig. 1
δ^*	displacement thickness
θ	momentum thickness
ξ	kinetic energy loss coefficient, see Eq. (2)

Subscript

CL	tip clearance
is	isentropic
m	mean
max	maximum
p	profile loss
s	secondary loss
t	total
t_e	trailing edge
1	inlet
2	exit

INTRODUCTION AND RESEARCH OBJECTIVES

The estimation of gas turbine performance has attracted extensive research efforts in the past thirty years. Knowledge gained in these efforts has resulted in vast improvements in turbine design technology and understanding of fundamental phenomena involved. Equally significant are the creation of many prediction methods for estimations of turbine performance from these studies. The performance prediction methods can generally be divided into two major categories - the overall stage methods and loss component methods.

The overall stage methods deliberately ignore the effects of turbine bladings and aerodynamics. The performance prediction is based on testing data from a number of turbines with similar characteristics. Thus, they are largely ad-hoc in nature and their use is somewhat limited. Typically the turbine stage efficiency is expressed as a function of either flow parameter and loading factor (Smith, 1965) or sprouting velocity and blade angles (Glassman, 1972). These methods have been very viable in performance prediction for turbines designed prior to 1970. Since then, gas turbine designs have evolved considerably. The stage loading has increased while aspect ratios have decreased. In addition, work distribution, blade stacking and blade-profile optimization have become important design parameters. The overall stage methods have lost most of their importance.

The loss component prediction methods use a basic concept that the sum of a number of individual loss components gives the total loss of an entire turbine stage. As a contrast to overall stage methods, knowledge of turbine flow characteristics and blading details is important here. This method initially defines important influence parameters which account for aerodynamic and geometric effects in a stage. These are followed by separate evaluations of individual loss associated with each parameter. Although such a prediction approach appears to be more accurate and acceptable, almost all of the methods developed are based on simplified models which involve critical assumptions. For example, all predictions are based on the mean-line velocity triangle known a priori, it hence assumes that the entire transport process undergone by the working fluid is represented by the velocity characteristics at midspan. Moreover, even with the mean-line assumption, systematic variations of a large number of parameters and accurate evaluations of each individual effects are virtually impossible. In reality, these effects are interactive and fundamentally inseparable. The complexity of actual systems is often far beyond the model's limits. To accurately predict a turbine performance thus requires certain corrections or modifications to the models.

One of the major concerns in our nation's space propulsion program is to improve the performance of space shuttle main engine (SSME). Among many components in SSME, the two-stage, unshrouded, axial turbine in the high-pressure fuel turbopump (HPFTP) has attracted great attention in the past. Nevertheless, continuing research efforts are greatly needed to understand the physical phenomena and to improve the turbine performance. NASA Marshall Space Flight Center (MSFC) is currently constructing a highly instrumented turbine test article (TTA) which is capable of facilitating extensive measurements on loss components and stage efficiency. One primary objective is to develop a viable means for accurate prediction of HPFTP turbine efficiency.

As a prelude to the up-coming testing with TTA, it is necessary to review and evaluate different models and correlations for turbine losses/efficiency currently available in the literature. Knowledge gained from such a study will provide a valuable baseline information to clarify future research directions with TTA. The work described here is primarily for this purpose. Here, the HPFTP turbine efficiency is assessed using the mean-line loss component method, as mentioned earlier. The prediction chooses correlations from previous studies and is considered to be suitable for HPFTP turbine. However, these

correlations are generally obtained from experiments with conditions substantially different from those in HPFTP turbine. The major difference includes blading geometry, loading conditions and fluid properties, and, therefore certain levels of deviation and uncertainty are expected.

Results obtained from the presently chosen correlations will be compared with corresponding results calculated from the NASA developed, MERIDL-TSONIC-BLAYER (MTB) method. The MERIDL and TSONIC are quasi three-dimensional codes for inviscid flowfields (Katsanis, 1969; Katsanis and McNally, 1977), and BLAYER is an integral solver for the viscous boundary layer (McNally, 1970). The MTB method is widely used for present day turbine design, largely because it is much less costly in computing time and technically simpler than a full three-dimensional, Navier-Stokes simulation. However, its accuracy remains uncertain, particularly for the predictions of flowfields. Povinelli (1985) has suggested that improvement in velocity prediction near blade trailing edge and the suction surface is required.

By definition, the total loss in a turbine stage is expressed in terms of total loss of dynamic pressure evaluated at the stage exit, i.e.

$$Y = (p_{t1} - p_{t2}) / (p_{t2} - p_2) \quad (1),$$

where Y is the total pressure loss coefficient, and p_{t1} , p_{t2} and p_2 represent the total pressures at stage inlet and exit and static pressure at exit, respectively. The loss component method assumes that the total pressure loss is the sum of profile, secondary and tip leakage losses, i.e. $Y = Y_p + Y_s + Y_{CL}$. For modern gas turbine involved significant blade cooling, e.g. film cooling, coolant loss is often considered. Another loss measure is defined as the loss of kinetic energy as compared to the isentropic condition, i.e.

$$\xi = (V_{2is}^2 - V_2^2) / V_2^2 \quad (2)$$

where ξ is the kinetic energy loss coefficient, and V_{2is} and V_2 are the velocities at exit under isentropic and actual situations, respectively. The major difference between these two coefficients lies their response to Mach number (M) variations. ξ is insensitive to Mach number variation; while Y increases with the magnitude of Mach number. The relation between these two loss coefficients can be expressed as,

$$Y / \xi = [1 + (r - 1) \cdot M^2 / 2]^{r/(r-1)} \quad (3)$$

where r is the ratio of specific heat at constant pressure to that at constant volume. It is clear that Y and ξ are practically equal for the low-Mach number flow, and it is generally the case for HPFTP turbine. The maximum value of Mach number in HPFTP turbine at full power level is approximately 0.4.

The following describes the loss components and reviews the corresponding correlations reported in the literature. To aid the description, The blade terminology is shown in Fig. 1.

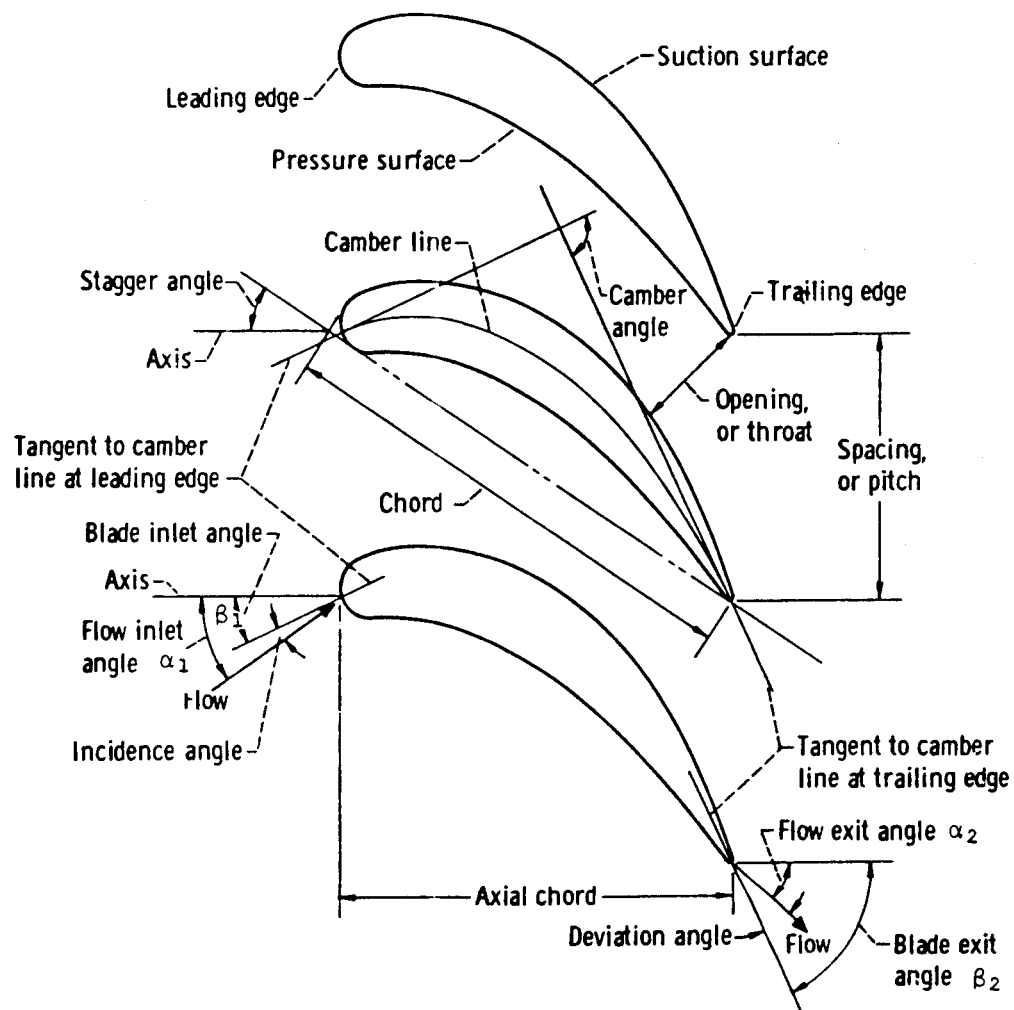


FIG. 1 BLADE TERMINOLOGY

DESCRIPTIONS OF LOSS COMPONENTS

1. PROFILE LOSS

The profile loss is composed of two parts. The first is the friction loss associated with boundary layer development on the blade surfaces, so-called the "basic profile loss". The second part represents the aerodynamic loss due to wake mixing near the blade trailing edge when the blade has a non-zero thickness. This leads to the name - "trailing edge loss". The phenomenon of this wake mixing is caused by sudden expansion of flow area and non-uniform pressure distribution in the spanwise (radial) direction at a stage exit.

1.1 Basic Profile Loss

Fundamentally, the basic profile loss is proportional to the magnitude of momentum thickness at the stage exit. Hence, it depends on stage geometry, i.e. blade shape and pitch-to-chord ratio, and Reynolds number. Since the turbine flow is largely turbulent, it is also affected by blade surface roughness. To a more detailed extent, curvature of blade surfaces influences the boundary layer developments on both pressure and suction sides of the blade. This is particularly important near the blade leading edge where a transition boundary layer exists. However, the curvature effects on the concave and convex sides of the blade are considered to offset each other, so the flat-plate approximation is a reasonable assumption.

Results concerning the basic profile loss prior to 1970 has been extensively reviewed by Denton (1973). In this review, Denton compared seven different correlations which, in addition to his own, include those well-known ones from studies by Ainley and Mathieson (1951), Stewart (1955), Traupel (1966), Balje and Binsley (1968), and Craig and Cox (1970). Among them three frequently mentioned correlations are listed as follows.

Ainley and Mathieson (1951)

$$Y_p = [Y_{\beta=0} + (\beta_1 / \alpha_2)^2 \cdot (Y_{\beta=\alpha_2} - Y_{\beta=0})] \cdot (t_{\max} / 0.2 c)^{\beta_1 / \alpha_2} \quad (4)$$

Stewart (1955)

$$Y_p = 1 - \left[\sin^2 \alpha_2 (1 - \delta^* - \delta_{te}^* - \theta^*)^2 / (1 - \delta^* - \delta_{te}^*)^2 + \cos^2 \alpha_2 (1 - \delta^* - \delta_{te}^*)^2 \right. \\ \left. \{ 1 + 2 \cos^2 \alpha_2 [(1 - \delta^* - \delta_{te}^*)^2 - (1 - \delta^* - \delta_{te}^* - \theta^*)] \} \right] \quad (5)$$

Traupel (1966)

$$\xi_p / (1 - \xi_p) = 0.006 (L/s) \cdot (V_m / V_2)^3 \cdot [1 + 3/4 (U/V_m) \cdot (s/L)]^2 / \cos \alpha_2 \quad (6)$$

The comparison made by Denton (1973) shows surprisingly large deviations among these correlations. This is mainly caused by the ad-hoc nature of each individual test and aerodynamic improvement toward modern blade design. The latter is evidenced by the fact that correlations proposed in earlier days tend to over-predict the loss for turbines developed later. Kacker and Okapuu (1982) suggest that the losses predicted by the correlation of Ainley and Mathieson (1951) should be multiplied by a factor of 2/3 to

account for the progress in blade design. In addition, all the studies aforementioned use stationary cascade models which neglect the rotational effects in actual turbine conditions. Dejc and Trojanovkij (1973) propose a correction factor to transform data with stationary cascades for use in actual turbine stages.

1.2 Trailing Edge Loss

According to conservation equations for continuity, momentum and energy, the trailing edge loss is the sum of two terms, i.e.

$$Y_{te} = [te / (1 - te)]^2 \cdot \cos^2 \alpha_2 + C_b \cdot te \quad (7)$$

The first term represents the loss caused by sudden expansion of flow area at the stage exit, and the second term is the additional mixing loss induced by the difference of pressures at blade base and average pressure across the exit span. C_b is the so-called "base pressure coefficient" which, in fact, bears a complex nature and varies strongly with blade geometry and flow conditions. The flow parameters affecting the value of C_b include the ratio of momentum thickness to blade trailing edge thickness and the Mach number of exit flows on both sides of the blade. Sieverding (1980) reported C_b data for flow exit-angle, $\alpha_2 = 60^\circ$ to 70° and $te/c = 0.06$ to 0.15 . His results confirm the strong dependency of C_b on parameters mentioned above.

Kacker and Okapuu (1982) observed a distinct difference of trailing edge loss between stators and impulse blades. They claimed that the difference in the thickness of boundary layer is the major cause of this phenomenon. Impulse blades, with their thick boundary layer, have lower (less negative) base pressure coefficients and thus have lower trailing edge loss. The correlation they proposed is

$$\xi_{te} = \xi_{te (\beta_F=0)} + |\beta_1 / \alpha_2| \cdot (\beta_1 / \alpha_2) \cdot [\xi_{te (\beta_F=\alpha_2)} - \xi_{te (\beta_F=0)}] \quad (8)$$

An important aspect pertaining to trailing edge loss, that is missing in the past research, is the effects of flow unsteadiness. The wake mixing inherits the flow characteristics of vortex shedding in various frequencies and blade interactions. These unsteady flow effects are generally non-linear with respect to time, and thereby have accumulative influence on the transports of momentum and energy. From the standpoint of dimensional analysis, the wake mixing and trailing edge loss are functions of many parameters, in particular Strouhal number and Reynolds number. Further studies in this regard should be emphasized.

2. SECONDARY LOSS

The secondary loss is due to the secondary flow in the blade passage. To gain a better understanding of the passage flow, it is usual to consider an ideal "primary" flow which may, for example, be two-dimensional potential flow. The difference between this primary flow and actual flow is then termed secondary flow. There are two mechanisms which have a major influence on the generation of secondary flows in turbine passages. The first mechanism is the effects of rotation and curvature, which cause the development of vortices in the streamwise direction. This is the similar mechanism of three-dimensional flow patterns existing in curved ducts, pipe bends and rotating channels. The vortices formed under this condition is the so-called "passage vortices" in turbine cascades. The second mechanism is the roll-up of endwall boundary layer in front of a blade. The flow

pattern, due to its particular shape, is called "horseshoe vortex," similar to its classical sense of cross flow over a wall-attached cylinder. The significance of horseshoe vortex in a turbine passage flow has been recognized only very recently. Fig. 2 displays an artistic sketch of passage secondary flow pattern originally presented by Klein (1966) and Langston (1980).

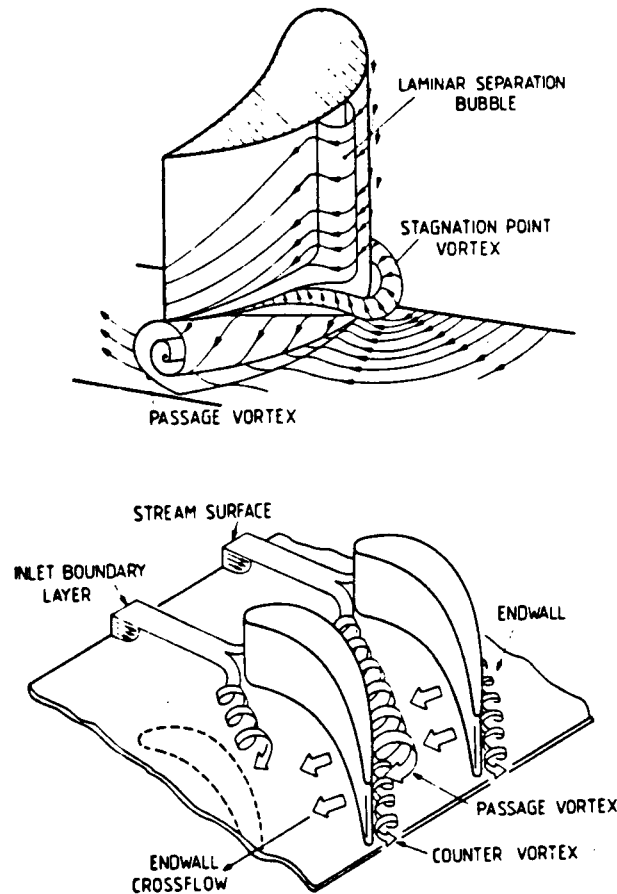


Fig. 2 SECONDARY FLOW MODEL BY KLEIN (1966) AND LANGSTON (1980)

The secondary loss, often amounting to nearly half of total stage loss, has been the subject of extensive research in the past. However, despite the large number of loss correlations available in the literature, it remains difficult and is risky to extrapolate the existing results for a generalized loss prediction. This is mainly due to the ad-hoc nature in passage geometry, blade aspect ratio, and flow condition for all the studies reported.

Secondary loss correlations prior to 1970 has been critically reviewed by Dunham (1970). His review lists 14 different correlations for incompressible flow. Secondary loss in these correlations is expressed by the total pressure loss coefficient defined as

Y_s = stagnation pressure loss / exit dynamic pressure.

In addition, all correlations are composed of a loading term, representing the effect of cascade loading or deflection and a length-scale ratio term. A typical form under this concept is expressed as

$$Y_s = (c/h) \cdot (\cos \alpha_2 / \cos \beta_1) \cdot f(\delta_1^*/c) \cdot Z \quad (9)$$

where, Z is the loading parameter given by Ainley and Mathieson (1951),

$$Z = (C_L/(s/c))^2 \cdot \cos^2 \alpha_2 / \cos^2 \alpha_m \quad (10)$$

where

$$\alpha_m = \tan^{-1} [(\tan \alpha_1 + \tan \alpha_2)/2] \quad (10a)$$

$f(\delta_1^*/c)$ represents the endwall boundary layer effect at the passage inlet.

Correlations presented later than 1970 still follows the same formula as Eq. (9); but with a more explicit form of the length-scale ratio term, i.e.

$$f(\delta_1^*/c) = C_1 + C_2 \cdot (\delta_1^*/c)^n \quad (11)$$

where C_1 , C_2 and n are constants, and their values vary with different studies. As an example,

$$\begin{array}{lll} C_1 = 0.0055, & C_2 = 0.078, & n = 0.5 \text{ (Dunham and Came, 1970),} \\ C_1 = 0.011, & C_2 = 0.294, & n = 1.0 \text{ (Morris and Hoare, 1975)} \\ C_1 = 0.034, & C_2 = 0 & n = 0 \text{ (Kacker and Okapuu, 1982)} \end{array}$$

In Eq. (11), the first term accounts for loss due to boundary layer growth on the endwall. The second term, a function of boundary layer thickness, represents the loss resulting from the horseshoe vortex formation near a blade leading edge.

3. TIP CLEARANCE LOSS

Aerodynamic loss due to flow leakage through the narrow gap between blade tip and adjacent outer casing represents a major efficiency penalty in a turbine rotor. The tip leakage flow is induced by the pressure difference between pressure and suction sides of a blade. The flow pattern is further complicated by the effect of relative wall movement with respect to the blade tip, which generates an additional secondary vortex, so-called the scraping vortex. Fig. 3 shows a schematic view of tip leakage flow. In modern turbine design, the tip leakage flow is controlled by maintaining close tolerance on tip clearance and/or geometric treatment of blade tip. Instead of plain tip, the latter commonly uses geometries such as winglet, squealer or groove tip.

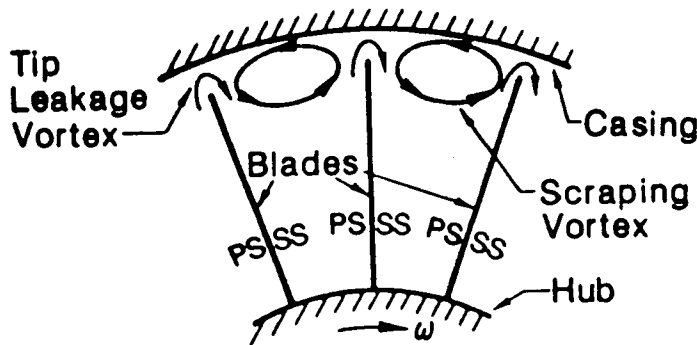


FIG. 3 TIP LEAKAGE FLOW

The fundamental mechanism of tip leakage flow that results in performance loss is not understood in detail. This, in part, is due to strong flow interaction and its effects on transport processes. Hence, for an unshrouded rotor stage, a clear demarcation between secondary and tip clearance losses is fundamentally impossible. This is particularly true for HPFTP turbine where the blade aspect ratios are small, nearly approximately unity. Under this condition, tip clearance loss is defined as the difference between total losses with and without clearance, i.e. $Y_{CL} / Y_{CL=0}$. Peacock (1982) has assessed the mechanism of tip clearance loss by considering the following four factors.

- (1). Pressure difference between pressure and suction surfaces of a blade - primarily an inviscid effect. This is the same mechanism leads to the tip vortices on an aircraft wing. The tip vortices progressively induces loss in lift of the wing section towards the wing tip.
- (2). Presence of boundary layer on turbine casing - a viscous effect. The boundary layer itself affects the tip region performance by inducing aforementioned scraping vortices which, in turn, results in a three-dimensional separation contributing to the loss of efficiency.
- (3). Relative movement between blade and boundary layer on the casing - primarily a viscous effect. For a turbine blade, this movement is in the opposite direction of tip leakage flow, thus it is in opposition to the mechanism stated in (1).
- (4). Size of clearance. For blades with large tip clearances, this effect overwhelms the combined mechanism of (1) to (3). This is significantly detrimental to the turbine efficiency due to blade unloading. For small clearance, the pressure force is balanced by the viscous force in the leakage flow, and the effects of (1) to (3) prevail.

For unshrouded, plain-tip, correlations of tip clearance loss prior to 1950 are expressed as linear functions of reduced flow area or mass flow rate in the presence of tip clearance. These correlations, largely based on steam turbine data, have been reviewed by Ainley and Mathieson in 1951. About the same time, Ainley (1951) proposes a correlation involving both clearance size and blade loading, i.e.

$$Y_{CL}/Y_{CL=0} = 0.5 \cdot (k/h) \cdot Z \quad (12)$$

where k and h are clearance size and blade height, respectively, and Z is the Ainley Loading factor. Hong (1962) later using an annular flow model reported a similar correlation, i.e.

$$Y_{CL}/Y_{CL=0} = 1 - 2.2 \cdot (k/h) \quad (13)$$

Dunham and Came (1970) modified Ainley's correlation (Eq. (12)) and suggested that conventional linear dependence of loss on clearance size should be replaced by a power law. Their correlation for a plain tip is

$$Y_{CL}/Y_{CL=0} = 0.47 \cdot (k/h)^{0.78} \cdot Z \quad (14)$$

More recently, Kacker and Okapuu (1982) have claimed that correlations developed earlier overpredicts the tip clearance loss of recently developed turbines. They proposed instead

$$Y_{CL}/Y_{CL=0} = 0.93 \cdot [k/(h \cos \alpha_2)] \cdot (R_T/R_m) \quad (15)$$

where R_T and R_m represent the tip radius and blade mean radius.

Note that all correlations up to date, including Eqs. (12)-(15), are derived based on simplified, non-rotating cascade models. Hence, they fail to include some essential effects, i.e. rotational force, scraping vortices, and relative wall-movement.

4. OTHER LOSSES

Depending on details of specific turbine conditions, there are other loss components which are significant and need to be addressed. Each of these (other) losses directly relates to one or more of those major losses as discussed in previous sections. In the literature, losses belonging to this category include the effects of blade thickness (Roelke and Hass, 1983), flow angles of incidence (Flagg, 1967), Mach number (Ainley and Mathieson, 1951; Kacker and Okapuu, 1982), Reynolds number (Traupel, 1966, 1977), and coolant flow (Hartsel, 1972; Tabakoff and Hamed, 1975; Ito, Eckert and Goldstein, 1980).

The influence of Reynolds number on the turbine performance is two-fold. The first effect lies on the development of boundary layer attached to both surfaces of the blade. Near-wall velocity profiles and momentum thickness is determined by the value of Reynolds number. The second effect relates to the blade surface roughness or finish. A turbulent boundary layer, sharply different from its laminar counterpart, is very sensitive to the surface condition. Thus, the degree of surface roughness along with Reynolds number determines the friction characteristics on a blade surface. This, in turn, will affect the basic profile loss to a certain extent.

The Mach number effect lies mainly in shock loss near a blade trailing edge. The effect varies strongly with blade thickness and passage shape. The shock loss is particularly important near the trailing section of a blade suction surface. Thus, Mach number is an additional parameter to be considered for the trailing edge loss. A further complication pertaining to trailing edge loss is due to the coolant ejection (Haas and Kofskey, 1977). The ejection of blade internal coolant often exists in modern gas turbines; however, all turbine blades in SSME do not have this feature.

The coolant loss in gas turbine, in a conventional sense, is referred to the turbine performance loss due to coolant injection of film cooling. Film cooling is one of the most effective means for blade cooling in modern gas turbine engines. In this case, coolant (compressed air in gas turbine) is directly injected into the hot mainstream, through either slots or discrete holes located on blade surfaces. As a result of momentum interaction between the injection and mainstream, a coolant film is formed covering the blade surface to be protected from hot-gas exposure. The direct impact of coolant injection is to thicken the boundary layer, so the profile loss increases.

In SSME HPFTP turbine, the coolant loss is fundamentally different from that of gas turbines as described above. First of all, current HPFTP design does not use film cooling or any other enhancement cooling to protect the blades. The coolant here is referred to the liquid hydrogen (LH_2), a cryogenic fluid, which is mainly used for cooling of rotating disks. Due to rotation, LH_2 gains radial component and flows outward. The coolant can bleed into the turbine passage through the gap (approximately 0.01") between rotor and stator. However, it is unclear that the flow passing through this gap is whether inward or outward, due mainly to the complex pressure distributions in HPFTP. The amount of flow rates are also uncertain. According to the present knowledge largely gained from numerical simulations, the combusted gas flows radially inward in the first stage (ahead of both stator and rotor) and LH_2 bleeds outward in the second stage. In either case, the predominant influence of this leakage flow lies in the region near a blade leading edge as it affects the structure of oncoming boundary layer before separation. The nature of leading-edge separation and horseshoe vortices existing in the blade-endwall section can vary significantly. As a result, the effects on the secondary loss is expected. So far, studies relating to this unique aspect of coolant loss in HPFTP have never been undertaken.

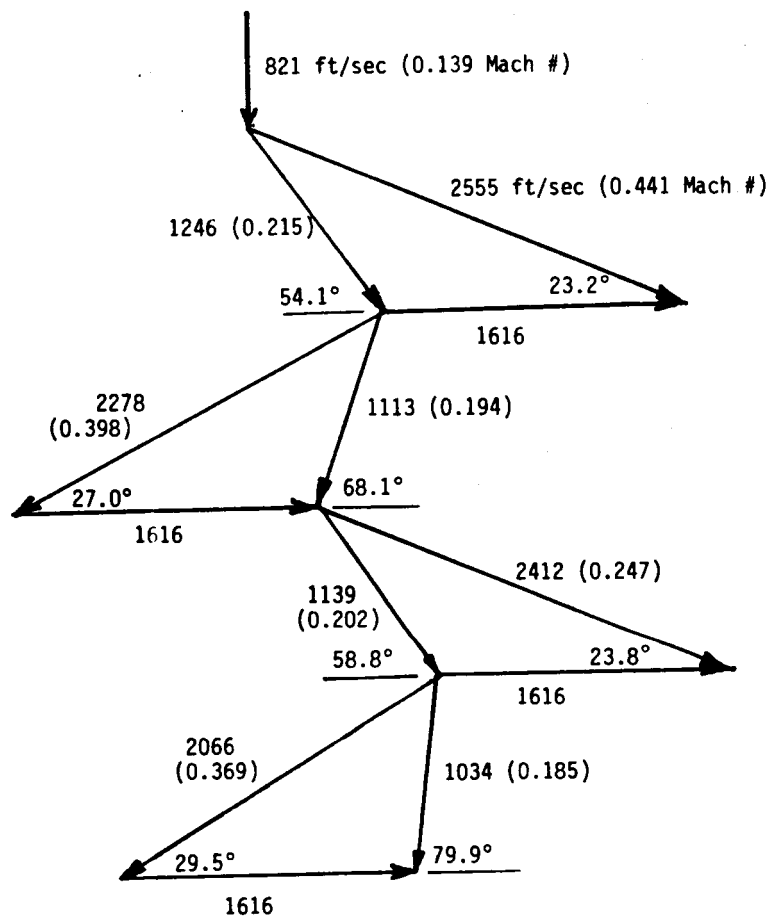
MEAN-LINE LOSS ESTIMATION

The present mean-line prediction uses the velocity triangle shown in Fig. 4. This full power-level velocity diagram was provided by Rocketdyne engine-balance calculation on May 21, 1987. Also listed in the figure is information pertaining to the specific turbine operation, e.g. pressure, temperature, mass flow rate, and rotating speed. The height of blade (h) is 0.884 and 0.984 inch for the first and second stage, respectively. The mean axial width is about 20% larger than the height for the first stage nozzle; nevertheless, this feature is completely reversed for the second stage rotor. The first stage rotor and second stage nozzle have a width-to-height ratio of nearly unity. The passage pitch-to-height ratio is approximately 0.88, 0.58, 0.86 and 0.55 for the first nozzle, first rotor, second nozzle and second rotor, respectively. The tip clearance used here is 0.019 and 0.016 inch respectively for the first and second stage rotor, and each of these amounts to 2 and 1.5% of the corresponding blade height.

To calculate the transport properties of working fluid in HPFTP, the present study uses correlations of real gas properties reported recently by Harloff (1987). These correlations include all important fluid properties in polynomial forms of temperature with a fixed gas mixture ratio of 86.87% H_2 and 13.13% H_2O . Using these polynomials removes the limitation of using air-equivalent conditions. Moreover, properties presented in polynomial forms are required as an input for loss predictions using the MERIDL-TSONIC-BLAYER (MTB) computer code. In the same study, Harloff has performed such a computation, and his results will be compared with the present ones.

Table 1 shows the loss-component breakdown and efficiency for each individual stage and entire turbine. Also shown in the parenthesis is the corresponding results from Harloff's (1987) MTB numerical computation. Mainly because of limited flow information, particularly the flow angles, the present mean-line prediction neglects the losses contributed by the trailing edge mixing, flow incidence, and effects of Reynolds and Mach number. Here, the estimation of basic profile loss uses the Stewart (1955) correlation, Eq. (5), since it bears many important fundamental features and is also used in the Harloff's simulation. The correlation used for secondary loss is the one developed by Kacker and Okapuu (1982), Eq. (8), which is considered to be appropriate for turbines designed later than 1970. The tip clearance loss uses the correlation by Dunham and Came (1970), Eq. (14), which involves power-law relationship, instead of the conventional linear one, with the length scale ratio term.

A comparison between the results from two studies shows that, except for the tip clearance loss, the present mean-line method gives higher values of loss prediction. In both studies, the greatest loss contribution comes from the profile loss in the stator, and tip clearance loss in the rotor. The tip clearance loss accounts for nearly one-half of the total rotor loss. The differences in profile and secondary losses are substantially large, often exceeding 100%. The entire, 2-stage, turbine efficiency is estimated to be 76.1%, compared to 81.6% reported by Harloff. The Rocketdyne design efficiency is cited as 79.1%. The present mean-line method appears to be the most conservative one, and it may overpredict the losses. Note that the velocity diagrams among all three studies are slightly different. However, the effects on the loss prediction results are considered to be very minor.



	Temperature R	Pressure PSIA
Inlet Total	1892	5525
1st Nozzle Exit Static	1829	4812
1st Rotor Exit Static	1794	4393
2nd Nozzle Exit Static	1749	3939
2nd Rotor Exit Static	1720	3648
Rotating Speed (RPM)	36353	
Power (HP)	72903	
Flow Rate (LB/SEC)	164.9	

FIG. 4. VELOCITY TRIANGLE

	1st Stator	1st Rotor	2nd Stator	2nd Rotor
1. Profile Losses Strwart (1965)	0.022 (0.014)	0.025 (0.015)	0.022 (0.014)	0.021 (0.016)
2. Secondary Losses Kacker & Okapuu (1982)	0.017 (0.012)	0.026 (0.010)	0.022 (0.011)	0.029 (0.011)
3. Tip Clearance Losses Dunham and Came (1970)		0.047 (0.063)		0.033 (0.047)
4. Incidence Losses	(0.001)	(0.006)	(0.001)	(0.014)
5. Stage Efficiency		0.867 (0.902)		0.878 (0.905)
6. Turbine Efficiency			0.761 (0.816)	

Rocketdyne Design Efficiency: 0.791

TABLE 1. LOSS BREAKDOWN

CONCLUSIONS AND RECOMMENDATIONS

An extensive literature review on the component loss prediction for unshrouded axial turbine has been performed in this study. All the correlations presently available are generally ad-hoc in nature and may be inappropriate for accurate prediction of turbine performance in SSME. Future testing with Turbine Testing Article at Marshall Space Flight Center should be directed towards collecting sufficient data base as well as gaining fundamental understanding of transport phenomena in turbine flow. The data base is required for validation of computer codes, and the fundamental information is needed for theoretical model developments. The models developed must be capable of providing physical insight into both momentum and energy transport, and also improve the performance predictions for other types of turbines. The following recommends future research directions which are important and need to be studied in greater depth.

1. Effects of Reynolds number. The primary concern lies on its influence on the boundary layer structure with various degrees of blade surface finish. This has direct influence on the basic profile loss. The combined and/or interactive effects between the surface roughness and curvature are also of interest.
2. Loss due to flow unsteadiness and turbulence. The wake shedding associated with periodic blade interaction is recently known to have significant influence on the time-averaged blade surface heat transfer. This has been a major issue in turbine research community for the past two or three years. The analogy between the energy transfer and momentum transfer implies that such a flow unsteadiness can induce an excessive loss in turbine aerodynamic performance. A viable prediction method in the future must incorporate this aspect.
3. Tip Clearance Loss. Based on the fact that the tip clearance loss amounts to nearly 50% of total rotor loss, correlations with much higher accuracy are desirable. The correlations currently available are fundamentally inaccurate, as they fail to include some predominant features in the system, e.g. relative wall movement, scraping vortices, and unloading near the blade trailing edge. Study recommended requires detailed pressure and flow measurements near the tip region. Ultimate research results should include recommendations of optimal tip geometries for future SSME turbines. In addition, the interaction between the tip clearance loss and secondary loss also needs to be emphasized, particularly for the present HPFTP turbine having low aspect ratios.
4. Coolant Loss. Investigation in this aspect must be directed to the special features of present HPFTP design, as discussed earlier. A rough estimation of the coolant velocity through the gap between the stator and rotor disks is approximately 150 ft/sec (nearly 5 to 10% of the passage mean flow velocity), that is sufficiently strong to affect the flow near the blade endwall region. Despite its thermal effects, it is certain that the secondary loss will be affected to a great extent. Due mainly to the complex nature involved, it would be desirable to initiate the study by conducting experiments in a non-rotating environment. Having established the stationary model, the rotating effects can then be enforced.

REFERENCES

- Ainley, D.G. and Mathieson, G.C.R., 1951, "A Method of Performance Estimation of Axial-Flow Turbines," ARC R&M 2974.
- Ainley D.G., 1951, "Estimation of the Change in Performance Characteristics of a Turbine resulting from Changes in the Gas Thermodynamics Properties," ARC R&M 2973.
- Balje, O.E. and Binsley, R.L., 1968, "Axial Turbine Performance Evaluation. Part B-Optimization With and Without Constraints," J. Engineering for Power, pp. 349-360.
- Craig, H.R.M. and Cox, H.J.A., 1971, "Performance Estimation of Axial Flow Turbines," Proc. Inst. Mech. Engrs., Vol. 185, pp. 407-424.
- Dejec, M.E. and Trojanovskij, B.M., 1973, "Untersuchung und Berechnung Axialer Turbinenstufen," Verlag Technik, Berlin.
- Denton, J.D., and Singh, U.K., 1979, "Time-Marching Methods for Turbomachinery Flow Calculation. Part I-Basic Principles and 20 Applications; Part II-Three-Dimensional Flows," Application of Numerical Methods to Flow Calculations in Turbomachine, VKI Lecture Series 1979-7.
- Denton, J.D., 1982, "An Improved Time-Marching Method for Turbomachinery Flow Calculation," ASME Paper 82-GT-239.
- Dunham, J., 1970, "A Review of Cascade Data on Secondary Losses in Turbines," J Mech. Engrg. Sci., Vol. 12, pp. 48-59.
- Dunham, J. and Came, P.M., 1970, "Improvement to the Ainley-Mathieson Method of Turbine Performance Prediction," ASME Trans., Series A, J. Engineering for Power, Vol. 92, No. 3, pp. 252-256.
- Fagg, E.E., 1967, "Analytical Procedure and Computer Program for Determining the Off-Design Performance of the Induced Flow," NASA CR-710.
- Glassman, A.J., 1972, "Turbine Design and Application," Vol. 1, NASA SP-290.
- Haas, J.E. and Kofskey, M.G., 1977, "Cold-Air Performance of a 12.766-Centimeter-Tip-Diameter Axial-Flow Cooled Turbine, II-Effect of Air Ejection on Turbine Performance," NASA TP-1018.
- Harloff, G.J., 1987, "Real Gas Properties and Space Shuttle Main Engine Fuel Turbine Performance Prediction," ASME Paper 87-GT-106
- Hartsel, J.E., 1972, "Prediction of Effects of Mass-Transfer Cooling on the Blade-Row Efficiency of Turbine Airfoils," AIAA Paper 72-11
- Hong, Y.S., 1966, "Axial Turbine Loss Analysis and Efficiency Prediction Method," Boeing Co. Rep. D4-3220.
- Ito, S., Eckert, E.R.G. and Goldstein, R.J., 1980, "Aerodynamic Loss in Gas Turbine Stage with Film Cooling," J. Engineering for Power, Vol. 102, pp. 964-970.

Kacker, S.C. and Okapuu, U., 1982, "A Mean Line Prediction Method for Axial Flow Turbine Efficiency," J. Engineering for Power, Vol. 104, pp. 111-119.

Katsanis, T., 1969, "FORTRAN Program for Calculating Transonic Velocities on a Blade-to-Blade Stream Surface of a Turbomachine," NASA TN D-5427.

Katsanis, T. and McNally, W.D., 1977, "Revised Fortran Program for Calculating Velocities and Streamlines on the Hub-Shroud Midchannel Stream Surface of an Axial-, Radial-, or Mixed-Flow Turbomachine or Annular Duct," I-User's Manual, NASA TN D-8430.

Klein, A., 1966, "Investigations of the Entry Boundary Layer on the Secondary Flows in the Bladings of Axial Turbines," BHRA T 1004.

Langston, L.S., 1980, "Crossflows in a Turbine Cascade Passage," J. Engineering for Power, Vol. 102, No. 4, pp. 866-874.

McNally, W.D., 1970, "Fortran Program for Calculating Compressible Laminar and Turbulent Boundary Layers in Arbitrary Pressure Gradients," NASA TN D-5681.

Morris, A.W.H. and Hoare, R.G., 1975, "Secondary Losses Measurements in a Cascade of Turbine Blades with Meridional Wall Profiling," ASME Paper 75-WA/GT-13.

Peacock, R.E., 1982, "A Review of Turbomachinery Tip Gap Effects, Part 1: Cascade," Int. J. Heat & Fluid Flow, Vol. 3, No. 4, pp. 185-193.

Povinelli, L.A., 1985, "Assessment of Three-Dimensional Inviscid Codes and Loss Calculations for Turbine Aerodynamic Computations," J. Engineering for Gas Turbine and Power, Vol. 107, pp. 265-276.

Roelke, R.J. and Haas, J.E., 1983, "The Effect of Rotor Blade Thickness and Surface Finish on the Performance of a Small Axial Flow Turbine," J. Engineering for Power, Vol. 105, pp. 377-382.

Smith, S.F., 1965, "A Simple Correlation of Turbine Efficiency," J. Royal Aeronautical Soc., Vol. 69, pp. 467-470.

Tabakoff, W. and Hamed, A., 1975, "Theoretical and Experimental Study of Flow Through Turbine Cascades with Coolant Flow Injection," AIAA Paper 75-843

Traupel, W. 1966, 1977, "Thermische Turbomaschinen," Springer Verlag Publication.

N89-21736

1988

NASA/ASIE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

X-RAY FLUORESCENCE SURFACE CONTAMINANT ANALYZER
A FEASIBILITY STUDY

Prepared by:

Hudson B. Eldridge

Academic Rank:

Associate Professor

University and Department:

University of
Central Arkansas
Physics Department

NASA/MSFC:

Laboratory:

Division:

Branch:

Materials and Processes
Engineering Physics
Physical Science

MSFC Colleague:

Ralph Carruth

Date:

August 12, 1988

Contract No:

NGT 01-002-099
University of Alabama

ABSTRACT

The bonding of liner material to the inner metal surfaces of solid rocket booster cases is adversely affected by minute amounts of impurities on the metal surface. Suitable non destructive methods currently used for detecting these surface contaminants do not provide the means of identifying their elemental composition. This report investigates the feasibility of using isotopic source excited energy dispersive x-ray fluorescence as a possible technique for elemental analysis of such contaminants.

A survey is made of the elemental compositions of both D6-ac steel, a common construction material for the booster cases, and Conoco HD-2 grease, a common surface contaminant. Source and detector choices that maximize signal to noise ratio in a "Recessed Source Geometry" are made. A Monte Carlo simulation is then made of the optimized device incorporating the latest available x-ray constants at the energy of the chosen source to determine the device's response to a D-6ac steel surface contaminated with Conoco HD-2 grease. An examination of the response for a grease layer of 100 angstroms, this is 1 milligram per square foot if the grease is spread uniformly over the square foot, should produce a count rate of about 1 count per second for a 1 curie iron 55 exciting source (0.82×10^{10} photons per second).

As a result of this optimization calculation, an iron 55 radioactive isotopic source excited x-ray fluorescence spectrometer that would mount on the existing robotic arm used for remote scanning of the inside of the large booster tank is proposed. It could operate in conjunction with the presently used contaminant detection devices. The resultant size and weight restrictions favor a cooled lithium drifted silicon semiconductor x-ray detector, and isotopic source. These same restrictions favor a Peltier cooling unit over the more common liquid nitrogen filled dewar.

Since calcium was chosen as an indicator for the grease, it is recommended that, prior to constructing the analyzer, spectra be accumulated from samples of D-6ac steel using conventional x-ray fluorescence equipment available at MSFC within the Chemistry Laboratories to ascertain the absence or presence of calcium in the steel.

ACKNOWLEDGMENTS

I am grateful for the opportunity that I have had to participate in the Summer Faculty Fellowship Program. My appointed task for the summer could not have been completed without the advice and assistance of my NASA colleague, Ralph Carruth, Chief, The Physical Sciences Branch, Engineering Physics Division, Materials and Processes Laboratory.

During the course of my summer, many helpful discussions were provided by Dr. Ray L. Gause of the Engineering Physics Division. Considerable advice and support by James Zwiener, and Roger Linton of that same Division was given throughout the execution of this study. The computations required during my summer visit were possible only because the expertise of Tom Schlenker, Tom Brister, Debbie Bowerman, and Debbie Hager of Grumman Aircraft Corporation was made so readily available.

I would like to thank Dr. Mike Freeman, whose attention to all details of the program's administration prior to, during, and after, made my visit to Marshall Space Flight Center an enjoyable, intellectually stimulating, and professionally rewarding experience. I am grateful to the National Aeronautics and Space Administration and the American Society for Engineering Education for their support of the Summer Faculty Fellowship Program.

INTRODUCTION

X-ray fluorescence spectroscopy is a technique for quantitative elemental determinations wherein the characteristic x-rays emitted by the excited atoms of various elements within the sample are analyzed as to energy and intensity. The energy of an emitted x-ray is related to the atomic number of the excited atom, hence the atomic species (Mosley's Law); while, the intensity of the emitted x-rays at a given energy is proportional to the number of atoms of a given species residing within the illuminated volume of the sample. There are a number of different methods of exciting the atomic species in a sample but the one that generally gives the most economical, most compact, and least weight assembly is the isotopic source.

In the isotopic source method of excitation, one uses a radioactive source that provides either x-rays or low energy gamma rays. To perform the analysis, one merely illuminates the sample with the photons from the source and then collects the fluorescent x-rays emitted by the sample with a suitable detector that will disperse these x-rays in energy. The spectra will in general be complicated by elastic and inelastic scattering of the incident x-rays by the sample as well as the fluorescent x-rays from the atomic species within the sample that are not of interest to the analysis.

By a judicious choice of the radioactive source used for excitation, one may limit the analysis to a selected range of the atomic species present. A detector may also be chosen that is most sensitive to a selected range of x-ray energy to further eliminate the contributions of the undesirable x-rays. Finally filters may be inserted in the incident beam, the scattered beam, or in both to reduce the background.

The availability of high resolution high efficiency solid state x-ray detectors, fast pulse processing electronics, and personal computers with adequate speed and memory capacity has provided the means to perform a complete elemental analysis of certain specimen. The technique's exploitation is limited by the availability of exciting sources of suitable energy and adequate intensity, interference by elements in the sample that are not of interest, and the count rate limits of the pulse handling electronics. With currently available components, analyses are routinely made in a matter of minutes to detection limits of 1 microgram (Rachetti and Wegscheider, 1986).

THE ANALYTE: CONOCO HD-2 GREASE

Conoco HD-2 Grease is composed (Law, 1987) principally of a light paraffin oil (70%), the calcium soap of mixed aliphatic sulfonic acids and alkylaryl sulfonic acids (15%), and the calcium soap of mixed aliphatic carboxylic acids (8%). It also contains inorganic calcium bases (~5%), calcium carbonate (2%), and diphenyl amine (0.3%). Based on the above mentioned work of Law, the elemental composition of HD-2 Grease as calculated for this work is shown in Table 1.

TABLE 1 - Elemental Composition of Conoco HD-2 Grease

Element	Weight Fraction
CALCIUM	.0067
CARBON	.00306
HYDROGEN	.00022
SULPHUR	.00002
SODIUM	.00008
CHLORINE	.0032
NITROGEN	.00004
SILICON	.000028
MANGANESE	.000008
IRON	.000054
MAGNESIUM	.000028
ALUMINUM	.000028
COPPER	.00002
ZINC	.000002

THE SUB-STRATA: D-6ac STEEL

Ladish D-6ac is quite similar in composition to AISI/SAE 4340 a member of the family of medium carbon low alloy ultrahigh-strength steels (Philip, 1980). Since the actual concentrations of the additives in D-6ac are not given, those for 4340 will be used with the following modifications: (1) the upper limit for each element's concentration will be used, (2) vanadium will be added at the same rate as nickel (3) the amount of nickel will be the same as 4340. With these modifications, the pertinent properties of the elements in D-6ac steel are presented in Table 2.

TABLE 2 - Elements In D-6ac Steel used in Background Calculation

Element	Weight Fraction	Molecular Weight	Atom Fraction	Density grams/cm ³
Carbon	.0043	12.01	.02	2.25
Manganese	.0080	54.94	.008	7.3
Silicon	.0035	28.09	.007	2.24
Chromium	.0090	52.01	.01	6.93
Nickel	.0200	58.71	.02	8.8
Molybdenum	.0030	95.95	.002	9.0
Iron	.9322	55.85	.9322	7.86
Vanadium	.0200	50.95	.02	5.6

ANALYSIS TECHNIQUE

Elements in the grease that were not also present in the steel could be used as an indicator of the presence of grease on the steel surface. In addition, if the concentration of the element was consistent within the grease, a quantitative measurement of the amount of that element could also be a quantitative measure of the amount of grease. Upon examination of Table 3, the "----" means that a concentration for that element was not reported, and the elements in it with regard to their x-ray characteristics, one finds that carbon, silicon, iron, and manganese which are in the grease are also in the steel. The energy of the K alpha line from the elements hydrogen, sulfur, sodium, chlorine, magnesium, nitrogen, and aluminum is so low that they would not be effectively detected if the system was operated in the atmosphere using a detector with a standard entrance window. For the application proposed in this study, the system would be operated in just that environment. Under such operating conditions, these x-rays would be severely absorbed within the intervening medium between sample and detector sensitive volume. The elements copper, and zinc have a K alpha excitation energy greater than iron. This means that, if a source were chosen with a high enough energy to excite these elements, this source would also excite the iron in the steel sub-strata. The fluorescent x-rays from the iron, because of the large amount of iron present, would interfere with the the detection of the much smaller amounts of these elements in the grease layer. The element calcium is left as the best possible candidate to serve as an indicator of the presence of the grease on the steel.

TABLE 3 - WEIGHT FRACTIONS OF ELEMENTS IN HD-2 GREASE AND D-6ac STEEL

ELEMENT	D6-ac STEEL	HD-2 GREASE
CARBON	.0043	.00306
MANGANESE	.0080	.000008
SILICON	.0035	.000028
CHROMIUM	.0090	.-----
NICKEL	.0200	.-----
MOLYBDENUM	.0030	.-----
IRON	.9322	.000054
VANADIUM	.0200	.-----
CALCIUM	.-----	.0067
HYDROGEN	.-----	.00022
SULPHUR	.-----	.00002
SODIUM	.-----	.00008
CHLORINE	.-----	.0032
NITROGEN	.-----	.00004
MAGNESIUM	.-----	.000028
ALUMINUM	.-----	.000028
COPPER	.-----	.000002
ZINC	.-----	.000002

The above discussion has also fixed the radioactive source that could best serve in the analyzer as iron 55. This isotope decays by electron capture to manganese 55 which then emits its characteristic x-rays as the electron cloud around the nucleus rearranges to accommodate the daughter nucleus. Of the resultant electromagnetic decays, 25 percent are radiative with 22.1 percent being at an energy of 5.9 kilo electron volts (KeV) and 2.9 percent being at an energy of 6.5 KeV. These x-rays are too low in energy to appreciably excite the iron in the steel and low enough to effectively excite the calcium in the grease. The K alpha excitation edge for calcium is 4.09 KeV.

There are a variety of designs for the geometrical arrangement of the source, sample, and detector to perform a x-ray fluorescence type of analysis. It is generally accepted that the "Recessed Source Geometry" provides the greater detection efficiency (Conde and dos Santos, 1985) for a proportional type counter and this geometry will be applied in this study for the Si(Li) detector. The geometrical arrangement for the recessed source geometry is shown in Figure 1. The radioactive face of the

SOURCE : Fe^{55}
DETECTOR : $\text{Si}(\text{Li})$

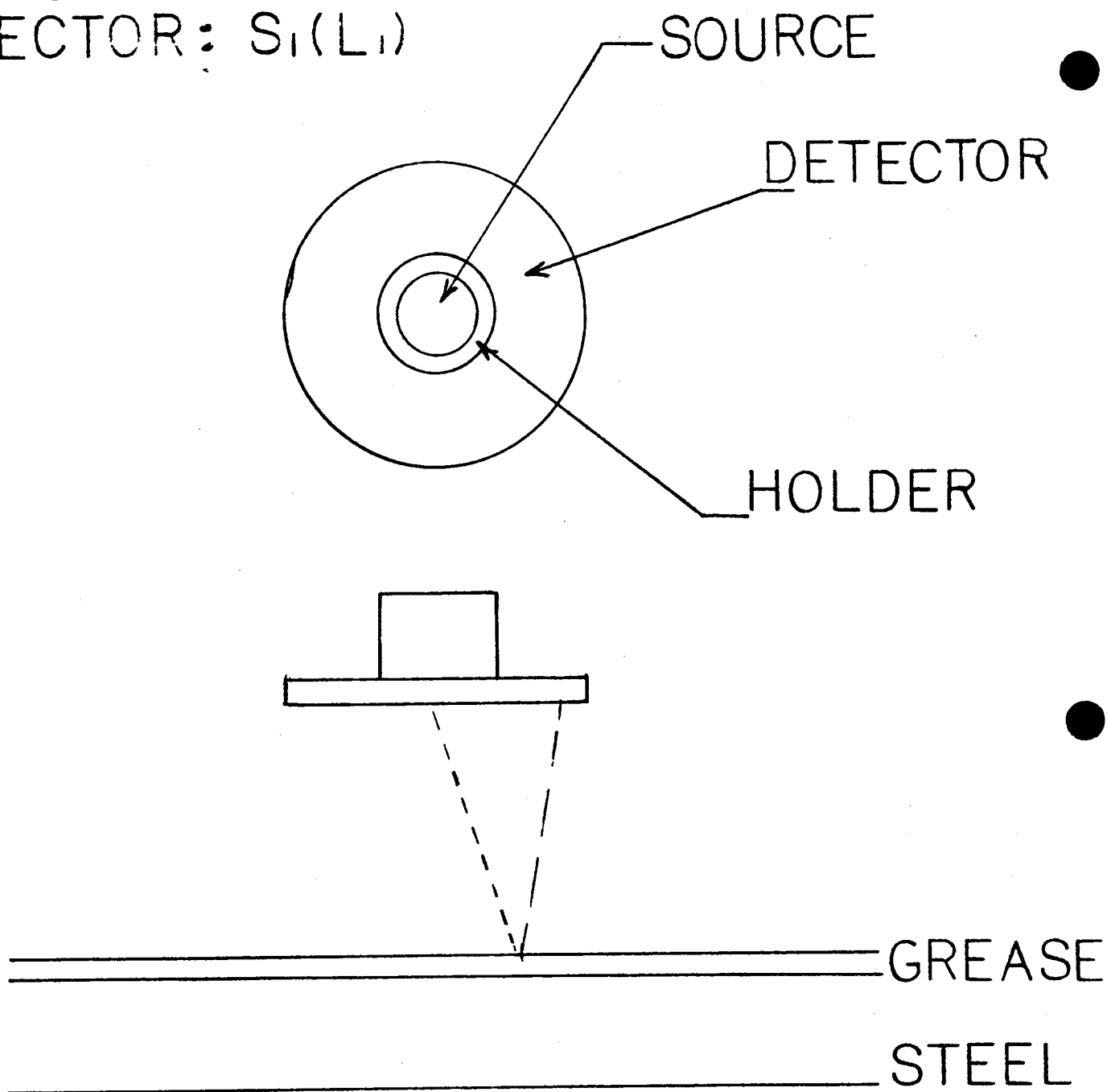


FIGURE I
X-RAY FLUORESCENCE
SURFACE CONTAMINANT
ANALYZER

source is flush to the face of and coaxial with its holder assembly whose diameter on that face is 4 mm greater than the source diameter. The source, with its holder, is then mounted flush to the face and coaxial at the center of the silicon wafer that is the sensitive volume of the x-ray detector. The radioactive face of the source is on the same side as the entrance face of the detector. An x-ray emitted from the iron 55 source that causes a calcium atom in the grease layer to fluoresce and the resultant fluorescent x-ray from the excited calcium atom is then absorbed within the detector's sensitive volume would provide the mechanism by which the device responded to the presence of grease on the steel. There would also be fluorescent x-rays impinging upon the detector from other atoms excited in both the grease layer and the underlying steel strata. These other fluorescent x-rays from the steel would be background and would serve to only degrade the signal to noise ratio for the experiment and prescribe the degree of energy resolution and pulse processing rate that the detector and ancillary electronics must have. Any other elements that were present could provide an additional measurement of the amount of grease present. It is possible that contaminants other than grease could be quantitatively measured on the surface, simultaneously, provided adequate energy resolution and counting rate capabilities were present in the detection system.

SIMULATION OF THE ANALYZER

The straightforward calculation of the response of the device shown schematically in Figure 1 involves the solution of a multidimensional integral equation wherein the integrand is not a simple function (Jenkins, Gould, and Gedcke, 1981). What is frequently done, instead of analytically solving the equation representing the response of the device, which of course would be generally preferred, is to compromise by performing a "Monte Carlo" or stochastic simulation of the process (Kalos and Whitlock, 1986), (Rubinstein, 1981), and (Hammersley and Handscomb, 1967). The first step in the simulation consists of dividing the physical process into a number of sequential events whose outcomes are governed by known probability distribution functions. The outcomes of these events are then determined by properly choosing from these probability distribution functions. The individual histories of the objects under study, in this case x-ray photons, are then compiled. These historical compilations are the desired solution which, in this case is, the response of the instrument in Figure 1.

For the particular case under study, the sequence of events starts with an circular area which emits photons of a definite energy in a random direction. The position that a photon is emitted from the surface is chosen from a distribution that is uniform, or constant, over the circular area and zero everywhere else. Once the position is chosen, the direction of emission is chosen from a distribution that is uniform in the polar angle over the region 0 to 90 degrees and uniform in the azimuthal angle from 0 to 360 degrees. Actually a photon emitted at a polar angle of 90 degrees would never reach the parallel surface of the sample, so a small "cut off" angle is introduced. This introduction has its attendant compromise to the accuracy but is necessary in a practical calculation. The net result is that photons of polar angles very near 90 degrees are discarded as failures at that point in their histories.

Once the position and direction of the emitted photon is decided, a path length or distance that it will travel before interacting with the medium is picked from the proper distribution. There are parameters in the distribution function of path lengths and the values of these parameters are determined by the nature of the medium and the energy of the photon.

At the interaction site, the type of atom that the photon interacts with is chosen. The probability of interacting with a

particular atom species is just the fractional number of that species in the sample. What type of interaction that occurs at that site is then chosen. From the interaction site, the original photon could emerge at a different direction, a different energy, or both. The original photon could also disappear and a different photon emerge. Each type of event would be picked in a random fashion from its proper distribution function.

In general, each photon would be traced on through the sample until it was absorbed or impacted the sensitive volume of the detector. The number of a particular energy photons that struck the detector's sensitive volume would be the response of the device to a given element for a certain source strength. When the history of a photon ends, another photon is started out from the source and its subsequent history is chronicled. After a very large number of these histories are compiled, in theory there should be an infinite number, the ratio of the number of a particular energy photon that strike the sensitive volume of the detector to the total number of photons that left the source, is the efficiency of the device for detecting a particular element in the sample.

RANDOM NUMBER GENERATOR

Central to the Monte Carlo method of solving integral equations is the availability of a sequence of truly random numbers. It is postulated that a Monte Carlo simulation will be an accurate representation of a physical process if the number sequence used in the simulation is truly random. However, to have a truly random sequence not only must the sequence be of infinite length but also each number must have an infinite number of digits. Finite amounts of time allotted a mortal and finite sized computer registers rule out the use of a truly random sequence in the solution of most practical problems. One must therefore compromise the accuracy by using a sequence of numbers of finite length, each of which have a finite number of digits. This so called pseudo random sequence approaches the ideal random sequence as both its length and the number of digits in the individual members becomes infinite.

A sequence of random numbers used in this simulation were generated using the GW-BASIC function RND(1). The function, RND called with the argument (1), started the sequence with the same member each time the program was run. To test the randomness of the number sequence generated by RND(1), a program was written to generate the sequence and then determine the frequency distribution of the numbers within the sequence. The frequency with which numbers, whose values lay within the interval, occurred for ten uniform intervals between 0 and 1 is shown in Table 4, for a sequence of numbers with 5,000 members. The percent deviation from the average for the frequency is as high as 10%. The data in Table 4 imply that a sequence with 5000 members is not an adequate representation of an infinite sequence.

TABLE 4 - Frequency of distribution of random numbers
in 5,000 member sequence; GW-BASIC RND(1)

Interval	Frequency	Deviation	Percent
0 - .1	447	-53	10.6
.1 - .2	496	-04	00.8
.2 - .3	503	+03	00.6
.3 - .4	514	+14	02.8
.4 - .5	524	+24	04.8
.5 - .6	515	+15	03.0
.6 - .7	470	-30	06.0
.7 - .8	498	-02	00.4
.8 - .9	531	+31	06.2
.9 - 1.0	502	+02	00.4

Table 5 shows the frequency distribution for a sequence of 50,000 members. The maximum percent deviation from the average for this sequence is about one tenth of Table 4 indicating that that this sequence is a much better approximation to the ideal random sequence. As a compromise between accuracy and computing time, the sequences used in this simulation were greater than 250,000 members in length. The effect of computer word size on the randomness of the sequence was not explored.

TABLE 5 - Frequency of distribution of random numbers
in 50,000 member sequence; GW-BASIC RND(1)

Interval	Frequency	Deviation	Percent
0 - .1	5025	+25	00.5
.1 - .2	5026	+26	00.5
.2 - .3	5053	+53	01.0
.3 - .4	5025	+25	00.5
.4 - .5	5057	+57	01.1
.5 - .6	5082	+82	01.6
.6 - .7	4989	+11	00.2
.7 - .8	4964	-36	00.7
.8 - .9	4948	-52	01.1
.9 - 1.0	4931	-69	01.4

UNIFORM SOURCE

To check the part of the simulation program that generates a circular surface that emits photons uniformly from its surface, a test was made of a source of radius 2 mm. A cartesian coordinate system was chosen with the origin at the center of the circular source. Each coordinate, X and Y, was divided into 10 equal intervals over the range -2 to +2. The frequency of occurrence of the coordinates (X,Y) of the point of emission of a photon lying within the area bounded by a given interval in X and Y was then computed. One quarter of the circular source is shown in Figure 2 where the source is superposed on an X,Y grid of the area intervals. For a square which lies wholly within the boundary defined by the circle, the relative frequency of a photon being emitted from this area is just the area of the square divided by the total area of the circle.

As an example, consider the small square bounded by $1.2 < Y < 1.6$ and $0 < X < 0.4$. the area of the square is $(0.4) \times (0.4)$ or .16 square millimeters. The area of the whole circle, whose radius is 2 mm., is 12.566... square millimeters. The ratio of the two, the fractional number of photons emitted from that area on the source is 0.0127... . Consider now a square that lies wholly outside the area of the circle. No photons should be emitted from this area since it is outside the source area so the frequency of occurrence of coordinates for the emission point of a photon should be 0. For the squares that lie partially inside and partially outside the circle, in Figure 2, the relative number of photons being emitted from a point with coordinates within the square, is just the area of the square that is within the circle divided by the total area of the circle. This number may be calculated exactly from the expression:

$$\frac{1}{4\pi} \int_{1.2}^{1.6} \int_{1.2}^{\sqrt{4-y^2}} dx dy$$

for the particular square defined by $1.2 < Y < 1.6$ and $1.2 < X < 1.6$.

Table 6 is a compilation of the two dimensional frequency distribution in X and Y for the discrete areas shown bounded by

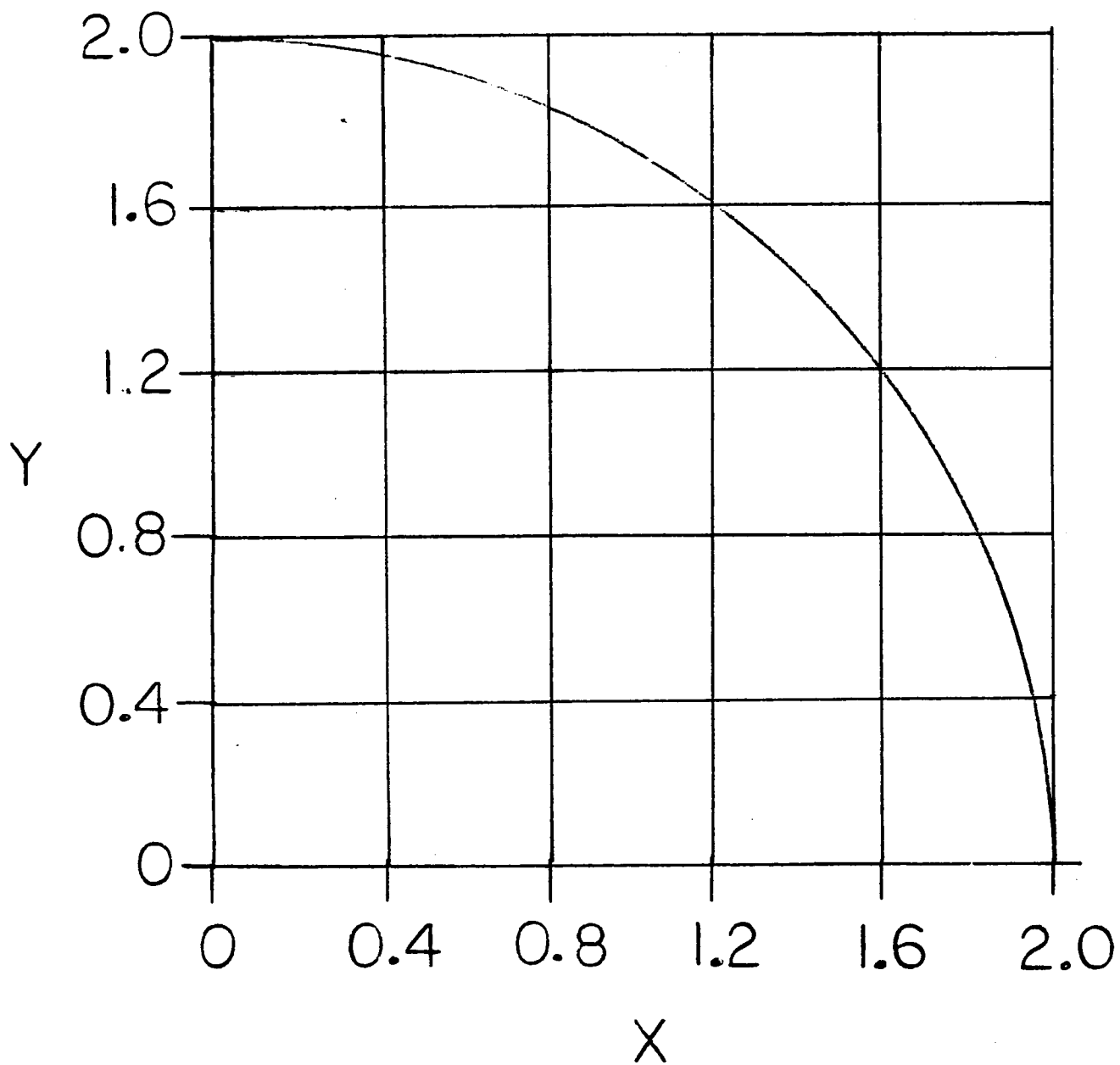


FIG. 2 PHOTON SOURCE
UNIFORMITY TEST

the Y intervals (vertically) and the X intervals (horizontally). By symmetry, a study of any one quarter of the circle is sufficient, actually there is also symmetry along a 45 degree line in the X,Y plane. The calculated frequencies, enclosed within parenthesis, are placed under the "MONTE CARLO" determined frequencies for comparison. Table 6 demonstrates that the simulated source shows the desired uniformity of emission over its area.

TABLE 6 - SOURCE UNIFORMITY TEST MONTE CARLO
(CALCULATED)
294,232 HISTORIES

INTERVAL		RELATIVE FREQUENCY				
Y	2.0-1.6	.012307 (.012513)	.009632 (.009948)	.003861 (.004138)	.000000	.000000
	1.6-1.2	.013031	.012725	.012517	.006947 (.006950)	.000000
	1.2-0.8	.012891 (.012733)	.012256	.012990	.012888	.004051 (.004138)
	0.8-0.4	.012738	.012847	.012901	.013051	.009435 (.009948)
	0.0-0.4	.012915	.012854	.012579	.012677	.012528 (.012513)
	INTERVAL	0.0-0.4	0.4-0.8	0.8-1.2	1.2-1.6	1.6-2.0
				X		

THE GEOMETRIC EFFICIENCY

At this stage in the simulation, we choose the area of the source and the height to operate the source from the sample surface that will give the optimum efficiency. The maximum size of the outer detector diameter is chosen as 30 millimeters by economic and resolution considerations. The wall thickness of the source holder was taken as 2 millimeters, a thickness that will be sufficient to protect the sensitive volume of the detector from direct illumination of the primary beam and still maximize the sensitive area of the detector available. This thickness is depicted, not to scale, in Figure 1 and labeled "HOLDER".

To study how the geometric efficiency varies with the above parameters, a computer program was written using the random number sequence generator and the uniform circular source studied previously. In this program, photons emitted by the source were assumed to have an infinite "path length" in the air between the source and sample and a 0 "path length" in the grease layer on the surface. This means that the photons suffer no attenuation in the intervening air but are absorbed immediately when they strike the grease. They then excite an atom in the grease and the resultant fluorescent x-ray is remitted isotropically. All the probabilities involved in the physical process at the interaction are taken as 1 to increase the efficiency, hence decrease the computing time, in this part of the study. Of these remitted photons, those that come off with a polar angle greater than 90 degrees are traced and, if one should end up in the sensitive volume of the detector, a successful outcome of an event would have occurred. Such a photon is depicted in a highly schematic fashion in Figure 2. The shorter dashes on the broken line indicates the incident photon from the source, higher energy or shorter wavelength; while the longer dashes on the broken line indicate the fluorescent x-ray emitted by an atom in the grease that was excited by the incident photon. The fluorescent x-ray would have a lower energy, or longer wavelength, than the incident photon; hence, the dashes in the line representing its path are longer.

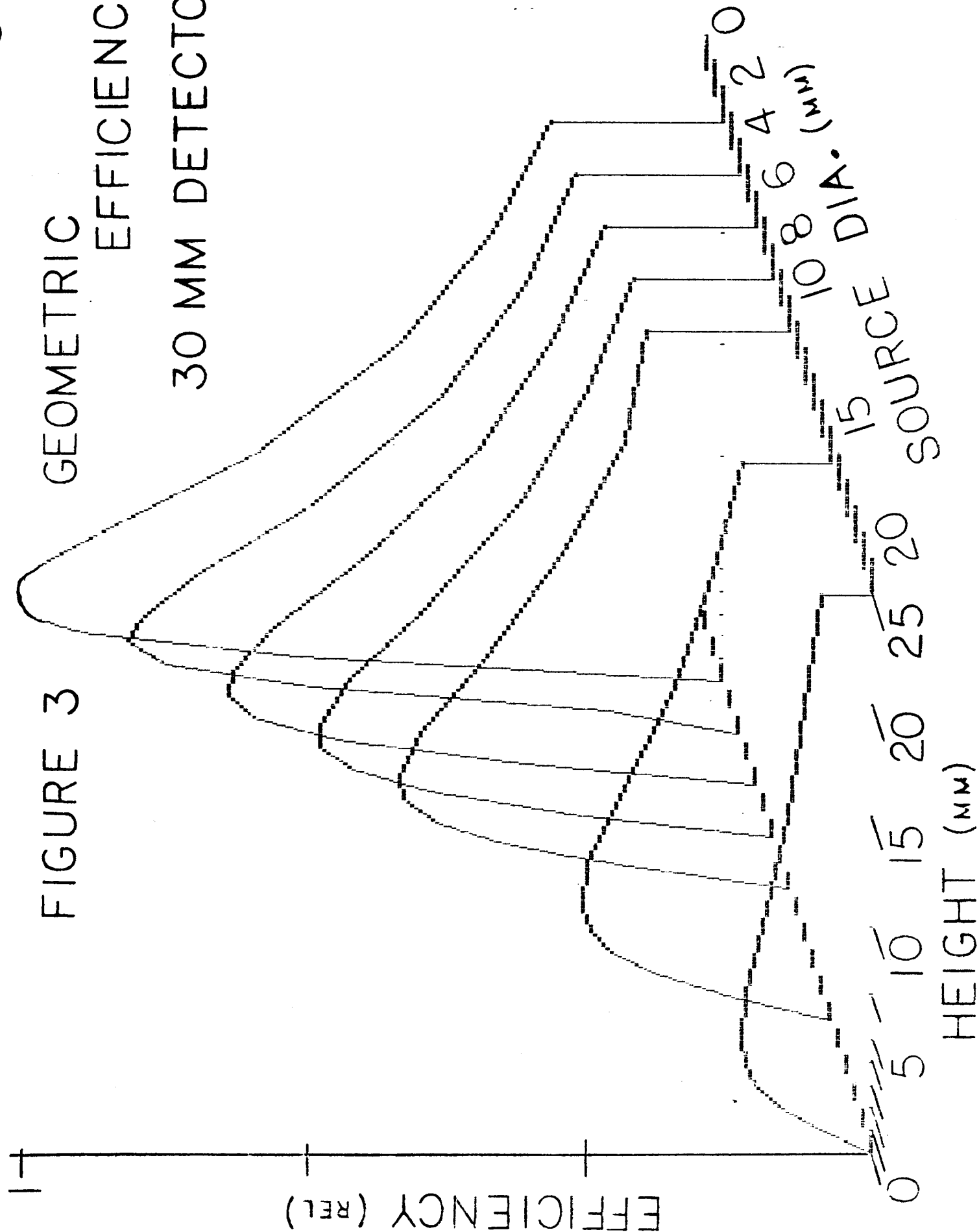
The results of the above study are shown in Figure 3. In this three dimensional plot, the height, in millimeters, of the source-detector assembly above the sample surface and the source diameter, in millimeters, are shown as the two perpendicular axes in the horizontal plane. The relative geometric efficiency of the device is plotted as the distance above that plane. The

FIGURE 3

GEOMETRIC

EFFICIENCY

30 MM DETECTOR



efficiency continues to increase as the source diameter is reduced, as would be expected, because the same radioactivity was assumed for the source regardless of its diameter. This situation is not physically realizable for the smallest source that can be constructed with a reasonable activity is about 4 millimeters in diameter. This size will be used in the rest of the study. At 4 mm source diameter, the geometric efficiency maximizes at a height above the sample surface of 4 mm so, in the rest of the study, the 4 mm diameter source will be placed 4 mm above the sample surface. The relative efficiency is shown in Figure 3. The magnitude of the efficiency, at the position of 4 mm above the surface for a 4 mm source, was 22.1%.

For the rest of the study, actual path length distribution functions will be used. Iron, however, will be substituted for D-6ac steel for purposes of path length calculation. The kind of atom impacted at each collision site will be selected according to the relative number of those kinds of atom per volume of the medium in which the collision occurred. The path length for the photons in air will, however, continue to be taken as infinite since such an approximation will introduce little error in the final results. In addition, second order collisions are neglected. This is reasonable at this stage since the flux available for second order collisions is about one millionth of the incident flux.

These features were added to the program that was used to compute the geometric efficiency and this new program was run for a number of cases where the thickness of the grease layer on the steel SUB STRATA was varied. The program generated the number of calcium K alpha x-rays originating in the grease layer that subsequently ended up in the sensitive volume of the detector. The number of K alpha x-rays from the impurity atoms in the steel that, after excitation by the primary beam, made their way to the detector sensitive volume, were also tabulated by the program. These values would be a lower estimate of the background that would be expected when using this device.

PATH LENGTH DISTRIBUTION

A flux of photons transiting a medium characterized by a mass absorption coefficient μ and a density ρ is attenuated according to the formula (Jenkins, Gould, and Gedcke, 1981):

$$I(x) = I(0) e^{-\mu \rho x}$$

where: $I(x)$ = the flux of at position x
 $I(0)$ = the flux at $x=0$
 μ = mass attenuation coefficient
 ρ = density of the medium

If the number of photons that successfully transit the medium of thickness x , $I(x)$, is divided by the number of photons that started out into the medium, $I(0)$, the probability, to within an arbitrary constant, of a photon having a path length in that medium is obtained. This is to say that the number of successful outcomes of an event (a photon surviving the transit) divided by the successful plus the unsuccessful outcomes of an event (a photon being absorbed) defines the probability of a given photon surviving a distance x in the medium. We want the total probability that a photon has a path length between 0 and infinity to be 1, that is all photons are absorbed in an infinite medium, so, to normalize the distribution we divide by the factor:

$$\int_0^{\infty} e^{-\mu \rho x} dx$$

The expression for the probability of a randomly emitted photon having a path length x is:

$$P(x) = \mu \rho e^{-\mu \rho x}$$

The CONOCO HD-2 grease that serves as a contaminant on the steel is principally a hydrocarbon compound. For computing the path distribution for the 5.9 KeV photons in the grease we will

use the published value for paraffin like hydrocarbons, (Weast, 1987) value of 10 centimeters squared per gram and the actual value of the density of the grease, (Conoco, 1985) as 1 gram per centimeter cubed. Substituting these values in the probability distribution $P(x)$, we obtain the expression for the path length distribution function for 5.9 KeV photons in HD-2 grease:

$$P(x) = e^{-x}$$

This expression must be solved for x , the path length, in terms of P , the probability of a particular x . This yields:

$$x = \ln \left[\frac{1}{P(x)} \right]$$

If P is now picked from a random distribution ranging from 0 through 1, x will be properly distributed.

To check these conjectures, 4,763,489 path lengths were generated with P in the above expression replaced with the random number generating function $RND(1)$. The frequency with the path lengths occurred within specific intervals of length are shown in Table 7 and labeled "MONTE CARLO". They are to be compared with the frequencies labeled "EXACT". The calculated or exact value for the frequency of occurrence of path lengths within an interval of length bounded by the values a and b is given by:

$$\int_a^b e^{-x} dx$$

TABLE 7 - 5.9 KeV PHOTONS IN GREASE
PATH LENGTH DISTRIBUTION
4,763,489 HISTORIES

INTERVAL (in mm)	MONTE CARLO	EXACT
0 TO 1	.63222	.63212
1 TO 2	.23246	.23254
2 TO 3	.08557	.08548
3 TO 4	.03147	.03147
4 TO 5	.01159	.01158
5 TO 6	.00423	.00426
6 TO 7	.00154	.00157
7 TO 8	.00058	.00058
8 TO 9	.00021	.00021
9 TO 10	.00008	.00008
10 TO INFINITY	.00004	.00004

Since the expression for the path length probability distribution function gives the expected frequency distribution of paths, the same expression, with the parameters μ and ρ for grease replaced with those for iron, is used to pick the path length in the steel SUB STRATA.

THE INTERACTION SITES

After a photon left the source along a selected polar direction, it suffers no attenuation until it contacts the grease layer. There it either suffers an interaction in the grease layer or not depending upon the random path length chosen for it and the value of its previously chosen polar angle. If its path length is less than the thickness of the grease layer divided by the cosine of the polar angle, it suffers a collision in the grease layer. If it suffers a collision in the grease layer, the probability that that atom is a calcium atom is given by the fractional number of atoms in the grease that are calcium.

Using the analysis of the grease discussed earlier (Law, 1987), the atom fraction of calcium is:

$$6.21 \times 10^{-5}.$$

The probability that a 5.9 KeV photon produces a K alpha x-ray when it is absorbed by a calcium atom is given by:

$$F = \omega_k \left(\frac{r_k - 1}{r_k} \right) \frac{\tau}{\mu}$$

where: ω_k = K alpha x-ray fluorescence yield
 r_k = K edge jump factor
 μ = Total mass attenuation coefficient
 τ = mass photoelectric absorption coefficient

For the element calcium and at 5.9 KeV, using recent experimental values for the physical quantities, (McMaster, Kerr Del Grande, Mallett, and Hubbell), the above expression yields:

$$0.146.$$

The probability that both events occur is the product of the two above probabilities:

$$(0.146) \times (6.21 \times 10^{-5}) = 9.067 \times 10^{-6}.$$

To make the program more efficient, with no sacrifice in accuracy, we will assign the K alpha photon that comes from such an encounter a "weight" of the above number.

If the photon transited the layer of grease without an interaction, it is restarted at the boundary between the grease and steel with the selection of a new path length using the same expression as was used for the grease but with the attenuation coefficient and density appropriate to iron, the major component of the steel.

At the end of the path length in the steel, a collision occurs in which the photon is absorbed. The type of atom in the steel that suffered the collision is selected by first generating a random number. The magnitude of the number determines the kind of atom that absorbed the photon according to Table 8.

TABLE 8 - METHOD OF CHOOSING KIND OF ATOM IN STEEL THAT ABSORBED A 5.9 KeV PHOTON THAT PASSED THROUGH THE GREASE LAYER

Random Number Value	Kind Of Atom	Energy K alpha X-Ray(KeV)	Considered After Collision	Why Not ?
0.000-0.020	CARBON	0.26	NO	ENERGY TOO LOW
0.020-0.028	MANGANESE	5.89	YES	
0.028-0.035	SILICON	1.74	YES	
0.035-0.045	CHROMIUM	5.41	YES	
0.045-0.065	NICKEL	7.47	NO	NOT EXCITED
0.065-0.067	MOLYBDENUM	17.44	NO	NOT EXCITED
0.067-0.087	VANADIUM	4.95	YES	
0.087-1.000	IRON	6.4	NO	NOT EXCITED

The pertinent physical constants related to the atoms in the steel that are considered beyond their first collision in the background determination and calcium, considered in the grease, are shown in Table 9. The values were taken from a compilation of experimental results, where such data existed, and the rest were taken from published calculations and extrapolations, (Bambynek, Crasemann, Fink, Freund, Mark, Swift, Price, and Rao). Where the total mass absorption coefficient was not available, the total

photoelectric absorption coefficient was used. This procedure is quite accurate at these low x-ray energies. The column headings are defined earlier where the K alpha emission probability was discussed.

TABLE 9 - X-RAY CONSTANTS AT 5.9 KeV IN IRON
FOR SELECTED IMPURITY ATOMS

Atom Kind	τ_P cm ⁻¹	μ_P cm ⁻¹	ω_K	r_K	F	Atom fraction	K alpha Probab.
CALCIUM	615.42	617.52	.163	9.112	.146	6.21×10^{-5}	9.067×10^{-6}
CHROMIUM	479.61	484.82	.280	8.779	.245	.0100	.00245
MANGANESE	560.00	560.88	.314	8.612	.277	.0080	.00222
SILICON	372.64	372.64	.047	10.442	.042	.0025	.00029
VANADIUM	3221.2	3221.2	.253	8.770	.224	.0200	.00448

Once a K alpha x-ray had been emitted by one of the minor elements in the steel, a path length and direction was chosen for it in the iron which, of course, depended upon its energy. Table 10 gives the energy of the K alpha x-rays for each element considered and its mass absorption coefficient in iron. This x-ray could be absorbed in the iron, which ended its history, or it could proceed out into the grease layer. In the grease layer it might also be absorbed, which ended its history. It could also escape from the grease layer and, if it did, and happened to also hit the detector's sensitive volume, then a fraction of a background event was recorded. The fraction recorded is given in the last column of Table 9.

TABLE 10 - MASS ABSORPTION COEFFICIENTS FOR BACKGROUND ELEMENTS
IN D-6ac STEEL

Element	K alpha Energy	$\mu\rho_{-1}$ cm ⁻¹	Mean Path centimeters
Manganese	5.89	675.96	0.01479
Silicon	1.74	5545.49	0.00180
Chromium	5.41	806.72	0.01240
Vanadium	4.95	867.62	0.01153

RESULTS, CONCLUSIONS, AND RECOMMENDATIONS

The final version of the GW-BASIC program that calculated the response of the device in Figure 1 to the calcium in the grease layer overlying a steel sub strata is listed as Appendix I. The program was run for several thickness' of the grease and the response for selected values of the grease layer thickness is tabulated in Table 11.

TABLE 11 - ANALYZER RESPONSE

	TIMES 100,000		
	5.9 KeV PHOTONS		
	4mm DIA. SOURCE		
	30mm DIA. DETECTOR		
	GREASE LAYER THICKNESS (A)		
	100	1000	10000
ELEMENT			
CALCIUM	.000015	.000047	.00037
MANGANESE	.326	.345	.306
SILICON	.0234	.0228	.0107
CHROMIUM	.477	.498	.445
VANADIUM	1.68	1.68	1.45

An examination of the response for a grease layer of 100 angstroms, this is 1 milligram per square foot if the grease is spread uniformly over the square foot, should produce a count rate of about 1 count per second in the calcium peak for a 1 curie iron 55 exciting source (0.82×10^{10} photons per second). This is a moderately intense source; however, 5.9 KeV photons can be attenuated quite easily so as not to present a health risk. An hours counting time would produce 3600 counts in the calcium peak for a precision, due to counting statistics, of 1.7%. It should be noted that the intensity of the calcium peak is not linearly related to the thickness of the grease layer.

The intensity of the background due to the trace elements in the steel are orders of magnitude greater than the calcium intensity. This should not present a problem in the operation of the device since Si(Li) detectors are available that provide an energy resolution of .15 KeV at an energy of 5.9 KeV.

As a result of this optimization calculation, an iron 55 radioactive isotopic source excited x-ray fluorescence spectrometer that would mount on the existing robotic arm used for remote scanning of the inside of the large booster tank is proposed. It could operate in conjunction with the presently used contaminant detection devices. The resultant size and weight restrictions favor a cooled lithium drifted silicon semiconductor x-ray detector, and isotopic source. These same restrictions favor a Peltier cooling unit over the more common liquid nitrogen filled dewar.

Electrical impulses, output from the detector, would be transported by coax cable from the detector to the modular electronics part of the spectrometer which could reside off the robotic arm in racks. There the impulses would be amplified and shaped by conventional components and fed into an analog to digital converter. The converters output would, in turn, be fed into a personal computer, such as an IBM PC. Both hardware and software, for a number of computers, are available to accumulate these electrical impulses into spectra and then provide analysis of resultant energy spectra of the x-rays impinging onto the detector.

It is also recommended that, prior to constructing an analyzer such as the one simulated in this study, spectra be accumulated from samples of D-6ac steel using conventional x-ray fluorescence equipment available at MSFC within the Chemistry Laboratories. These spectra should then be scanned for all peaks present and their relative intensities determined. The presence of calcium in this steel can then be determined and, if present above the limits of detection, the amount and consistency from sample to sample should be studied. The amount of calcium in the steel as well as its intra sample consistency, as compared to that of the grease, will indicate the feasibility of the application of x-ray fluorescence to this contaminant determination.

REFERENCES

- Law, R.D., "The Compositional Analysis of HD-2 Grease"
LSDI-88-7;TWR-16896, Morton Thiokol, Inc., Wasatch
Operations, PO Box 524, Brigham City, Utah, (1987).
- Philip, T.V., "Ultrahigh-Strength Steels", ASM Metals Handbook
9th Edition, Vol. 1, (1980).
- Conde, C.A.N and dos Santos, J.M.F., "The Recessed Source
Geometry For Source Excited X-Ray Fluorescence Analysis",
Advances in X-Ray Analysis, 29, 545-550, (1985).
- Jenkins, R., Gould, R.W., and Gedcke, D., Quantitative X-Ray
Spectrometry, Marcel Dekker, Inc., New York, (1981)
- McMaster, W.H., Kerr Del Grande, N., Mallett, J.H., and Hubbell,
J.H. "Compilation of X-Ray Cross Sections", UCRL - 50174,
TID-4500, UC - 34 Physics, Lawrence Radiation Laboratory,
University of California, Livermore, California.
- Bambynek, W., Crasemann, B., Fink, R.W., Freund, H.U., Mark, H.,
Swift, C.D., Price, R.E., Rao, P.V., "X-Ray Fluorescence
Yields, Auger, and Coster-Kronig Transition Probabilities"
Reviews of Modern Physics Vol. 44 Number 4., 716-814,
(1972).
- Conoco, Inc., Material Safety Data Sheet, P.O. Box 1267, Ponca
City, Oklahoma, (1985).
- Weast, R.C., Editor, Handbook of Chemistry and Physics, E136,
CRC Press, Boca Raton, Florida, (1987).
- Kalos, M.H. and Whitlock, P.A., Monte Carlo Methods, John Wiley
& Sons, New York, (1986).
- Rubinstein, R.Y., Simulation and the Monte Carlo Method, John
Wiley & Sons, New York, (1981).
- Hammersley, J.M. and Handscomb, D.C., Les Methodes de Monte
Carlo, Dunod, Paris, (1967).
- Rachetti, A. and Wegscheider, W., "Background Intensities and
Their Utilization in Quantitative Analyses by Monochromati-
cally Excited Energy-Dispersive X-Ray Fluorescence", Advan-
ces in X-Ray Analyses Vol. 30 pp 143, Plenum, N.Y., (1986).

APPENDIX I

GW-BASIC PROGRAM THAT SIMULATES AN X-RAY FLUORESCENCE SURFACE CONTAMINANT ANALYZER

```

100 REM COAXIAL SOURCE ANNULAR DETECTOR RESPONSE TO GREASE ON STEEL
110 DIM ELEMENT$(10),PROB(10),PATH(10),FLUX(10)
120 OPEN "A:ELEMENT.DAT" FOR INPUT AS #1
130 FOR I=1 TO 9:INPUT#1,ELEMENT$(I),PROB(I),PATH(I):NEXT I
140 CLOSE #1
150 CLS:LOCATE 5,10:INPUT"DIAMETER OF SOURCE (in millimeters)";S
160 R=S/2
170 LOCATE 7,10:INPUT"DIAMETER OF SOURCE CONTAINER (in millimeters)";C
180 C2=C/2
190 LOCATE 9,10:INPUT"DIAMETER OF SENSITIVE AREA OF DETECTOR (mm)";D
200 D2=D/2
210 LOCATE 11,10:INPUT"HEIGHT OF SOURCE-DETECTOR PLANE FROM SAMPLE
(mm)";H
220 LOCATE 13,10:INPUT"THICKNESS OF GREASE LAYER IN ANGSTROMS";T
230 T=T*.0000001 'CONVERT THICKNESS OF GREASE LAYER TO MILLIMETERS
250 J=1'LET THE SOURCE EMIT THE FIRST PHOTON
260 X=R*RND(1) 'POSSIBLE X COORDINATE SOURCE EMITTS PHOTON
270 Y=R*RND(1) 'POSSIBLE Y COORDINATE SOURCE EMITTS PHOTON
280 IF(X^2+Y^2)>4 GOTO 260 'GOOD POSITION OF PHOTON EMISSION FROM
SOURCE??
290 SIGN=RND(1) 'EQUAL LIKELYHOOD OF +,- FOR X
300 IF SIGN<.5 THEN X=-X
310 SIGN=RND(1) 'EQUAL LIKELYHOOD OF +,- FOR Y
320 IF SIGN<.5 THEN Y=-Y
330 TH=1.5707963#*RND(1) 'RANDOM THETA FOR DIRECTION OF EMITTED PHOTON
340 IF TH>1.5 GOTO 600 'EMITTED ALMOST PARALLEL TO SURFACE
350 PHI=6.2831853#*RND(1) 'RANDOM PHI FOR DIRECTION OF EMITTED PHOTON
360 L=H/COS(TH)'SAMPLE PLANE IS H(mm) ABOVE SOURCE SURFACE
370 X=X+L*SIN(TH)*COS(PHI)'X COORDINATE PHOTON STRIKES SAMPLE SURFACE
380 Y=Y+L*SIN(TH)*SIN(PHI)'Y COORDINATE PHOTON STRIKES SAMPLE SURFACE
385 K=1 'SET ELEMENT PARAMETERS FOR CALCIUM (IN THE GREASE)
390 LZ=RND(1) 'PICK RANDOM NUMBER FOR PATH LENGTH IN GREASE
400 IF LZ<.0000001 GOTO 600 'PATH TOO LONG; WILL CAUSE PROBLEMS; TOSS
410 PL=LOG(1/LZ)'PICK PATH LENGTH FROM PROPER DISTRIBUTION
420 IF PL>T/COS(TH) GOTO 700 'PHOTON PASSED THROUGH THE GREASE LAYER
430 X=X+PL*SIN(TH)*COS(PHI)
440 Y=Y+PL*SIN(TH)*SIN(PHI)
450 TH=3.1415927#*RND(1)'PHOTON INTERACTS IN TARGET ISOTROPIC SCATTER
460 IF TH<=1.6 GOTO 600 'MUST BE BACKSCATTERED TO HIT DETECTOR
470 PHI=6.2831853#*RND(1)'ALSO AZIMUTHAL ANGLE ISOTROPIC

```

```

480 L=-H/COS(TH)'DETECTOR IS H(mm) BELOW PLANE OF TARGET
490 X=X+L*SIN(TH)*COS(PHI)'X COORDINATE WHERE PHOTON HITS DETECTOR
500 Y=Y+L*SIN(TH)*SIN(PHI)'Y COORDINATE WHERE PHOTON HITS DETECTOR
510 IF(X^2+Y^2)<=C2^2 GOTO 600'PHOTON HITS SOURCE, CONTAINER INSTEAD OF
DETECTOR
520 IF(X^2+Y^2)>D2^2 GOTO 600 'PHOTON WIDE OF DETECTOR SENSITIVE VOLUME
530 FLUX(K)=FLUX(K)+PROB(K) 'ATOM FRACTION ELEMENT(K) * PROBABILITY Ka
X-RAY
600 A$=INKEY$
610 IF A$<>" " THEN J=J+1:GOTO 260
620 LPRINT S"(mm) ACTIVE DIAMETER SOURCE IN A";C;"(mm) DIAMETER
CONTAINER IN A";D;"(mm) DIAMETER DETECTOR";H;"(mm) ABOVE THE
SAMPLE";T;"(mm) OF GREASE"
630 LPRINT:LPRINT:LPRINT
640 LPRINT"NUMBER OF PHOTONS EMITTED";J:LPRINT:LPRINT
650 FOR I=1 TO 9
655 LPRINT ELEMENT$(I),PROB(I),PATH(I),FLUX(I)
660 NEXT I
695 END
700 K=9 'SET ELEMENT PARAMETERS FOR IRON
710 LZ=RND(1) 'RANDOM NUMBER FOR PATH LENGRH IN IRON
720 IF LZ<.0000001 GOTO 600 'IF TOO SMALL; CAUSES PROBLEM WITH LOG
FUNCTION
730 PI=PATH(K)*LOG(1/LZ) 'PATH LENGTH OF PHOTON IN IRON
740 Z=PI*COS(TH)+H+T 'UPDATE COORDINATES
750 X=X+PI*SIN(TH)*COS(PHI) 'POSITION FOR PHOTON THAT WENT THROUGH
GREASE
760 Y=Y+PI*SIN(TH)*SIN(PHI)
1000 ATOM=RND(1)'RANDOM NUMBER TO PICK TYPE OF ATOM HIT BY PHOTON IN
STEEL
1010 IF (ATOM>=0)*(ATOM<.02) THEN K=2:GOTO 600'HIT CARBON ATOM; DON'T
COUNT TOO LOW ENERGY
1020 IF (ATOM>=.02)*(ATOM<.028) THEN K=3:GOTO 1090'HIT A MANGANESE ATOM
1030 IF (ATOM>=.028)*(ATOM<.035) THEN K=4:GOTO 1090 'HIT A SILICON ATOM
1040 IF (ATOM>=.035)*(ATOM<.045) THEN K=5:GOTO 1090'HIT A CHROMIUM ATOM
1050 IF (ATOM>=.045)*(ATOM<.065) THEN K=6:GOTO 1090 'HIT A NICKEL ATOM
1060 IF (ATOM>=.065)*(ATOM<.067) THEN K=7:GOTO 1090 'HIT A MOLYBDENUM
ATOM
1070 IF (ATOM>=.067)*(ATOM<.087) THEN K=8:GOTO 1090'HIT A VANADIUM ATOM
1080 GOTO 600 'THESE ARE IRON INTERACTIONS; CAN'T EXCITE THE IRON Ka X-
RAYS
1090 TH=3.1415927#*RND(1) 'ISOTROPIC EMISSION OF Ka X-RAY
1100 IF TH<1.6 GOTO 600 'MUST BACKSCATTER TO HIT DETECTOR
1110 PHI=6.2831853#*RND(1)
1120 LZ=RND(1)
1130 IF LZ<.0000001 GOTO 1120

```

```
1140 PL=PATH(K)*LOG(1/LZ) 'PATH LENGTH IN IRON FOR Ka X-RAY OF Kth  
ELEMENT  
1150 Z=Z+PL*COS(TH)  
1160 IF Z>H GOTO 600 'PHOTON DID NOT GET OUT OF IRON  
1170 L=-Z/COS(TH)  
1180 GOTO 490  
1190 END
```

N89-21737

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

INTELLIGENT DATA REDUCTION FOR AUTONOMOUS POWER SYSTEMS

Prepared by:	Stephen A. Floyd
Academic Rank:	Assistant Professor
University and Department:	The University of Alabama - in Huntsville, School of Administrative Science Department of MIS/MSC
MSFC Colleague:	David J. Weeks
Date:	September 7, 1988
Contract No.	NGT-01-002-099 The University of Alabama

INTELLIGENT DATA REDUCTION
FOR AUTONOMOUS POWER SYSTEM MONITORING

by

Stephen A. Floyd
Assistant Professor of MIS
College of Administrative Science
University of Alabama - Huntsville
Huntsville, Alabama

ABSTRACT

Since 1984 Marshall Space Flight Center has been actively engaged in research and development concerning autonomous power systems. Much of the work in this domain has dealt with the development and application of knowledge-based or expert systems to perform tasks previously accomplished only through intensive human involvement. One such task is the health status monitoring of electrical power systems. Such monitoring is a manpower intensive task which is vital to mission success. The Hubble Space Telescope testbed and its associated Nickle Cadmium Battery Expert System (NICBES) have been designated as the system on which the initial proof of concept for intelligent power system monitoring will be established.

The key function performed by an engineer engaged in system monitoring is to analyze the raw telemetry data and identify from the whole only those elements which can be considered "significant." This function requires engineering expertise on the functionality of the system, the mode of operation and the efficient and effective reading of the telemetry data. Application of this expertise to extract the significant components of the data is referred to as data reduction. Such a function possesses characteristics which make it a prime candidate for the application of knowledge-based systems' technologies. This paper investigates such application and offers recommendations for the development of "intelligent" data reduction systems.

ACKNOWLEDGEMENTS

The author wishes to express sincere appreciation to all those involved in the NASA/ASEE summer faculty program. Specifically I wish to thank Dr. Mike Freeman, Ms. Ernestine Cothran, Dina Engler and Missy Dunn for their professional administration of the program.

To all the employees of Marshall's Electrical Power Branch who made my summer a very enjoyable and rewarding one, I express my sincere gratitude. In particular I wish to express my appreciation to Mr. Dave Weeks who sponsored me in the laboratory and also to Randy Baggett, Louis Lollar, Norma Dugal-Whitehead and Bryan Walls with whom I interfaced on a regular basis. To Alex Bykat, another Power Branch Summer Fellow with whom I shared an office, I extend my gratitude for the many work related and professional discussions in which we engaged. I hope he found the companionship as enlightening and enjoyable as I did. Appreciation is also extended to all the Branch and Laboratory personnel for their friendly and cooperative spirit.

Finally, I wish to thank my wife, Beth, who was kind enough to offer her word processing expertise to type in the bulk of the final report.

INTRODUCTION

As the exploration of space continues missions become much more complex and much longer in duration. Future missions such as Space Station, spaced-based radar, communication and surveillance satellites, strategic defense initiative (SDI) systems and military aircraft will thus require more sophisticated and intricate electrical power systems (EPS) [8], [37]. Space power is an extremely precious resource. The fact that almost every subsystem, especially those that support the human elements for manned missions, is dependent on power plus the fact that space power has historically cost about \$1000.00/KWH versus \$.05 per terrestrial KWH has placed space power high on NASA's priority list of research efforts. As was learned from Skylab, for which 15-18 ground support personnel were required to augment extensive crew involvement for an 8KW system, a major effort had to be directed toward autonomously managed electrical power systems.

Automating activities ordinarily performed by humans was seen as the primary means of reducing both airborne and ground support efforts and costs [9],[16],[36],[39]. Additionally, more fully autonomous power systems (as well as other subsystems) will be a necessity for deeper unmanned exploration of space where missions will require decisions and actions in "real-time". The time lags incurred with data transmission and remote intervention will not be acceptable in allocating and protecting the precious electrical power resources. In 1978, therefore, the Office of Aeronautics and Space Technology at NASA Headquarters directed NASA to undertake efforts towards accomplishing such autonomy. Since that directive various NASA efforts in conjunction with several contractors (including Martin Marietta, Rockwell/Rocketdyne, Boeing, TRW, Hughes and Ford Aerospace) and Universities (among them Auburn, University of Tennessee, Tennessee Tech., University of Alabama-Huntsville, Vanderbilt and Carnegie Mellon) have made much progress in the realm of space power automation.

It was realized early on in power system investigations that autonomous systems would require a certain amount of embedded intelligence to supplement the already proven more conventional computer approaches [23]. Thus much of the current research effort is focussed on artificial intelligence techniques, namely application of expert and knowledge-based systems. The term "expert system" (ES) refers to a software system which performs a complex, well defined task using the same input information and problem solving strategies as a human expert. Additionally, an expert system possesses the capability to make accessible to the user the reasoning logic it uses to perform the task. It is implied that the expertise captured by such a system has its origins in the experience that one or more humans have accumulated while performing a given problem solving task. The term "knowledge-based system" (KBS) refers to a software system much like an

expert system but which implements a body of problem solving knowledge which may come from any of several sources including text books, humans (in the form of expertise or more general experiential knowledge) or others.

It is important in the domain of space power system applications to draw the distinction between these two types of systems. The reason for this is that this is a very young domain and "experts" with experience managing space power systems do not exist. However, the experience of humans working in this arena coupled with more general knowledge about power subsystems and components make it possible to develop what for the purposes of this paper will be referred to as knowledge-based expert systems (KBES).

Though few doubt the important role that KBES approaches will play in space power automation the domain is one which offers more complex challenges than those to which the technology has already been successfully applied. One of the approaches to overcoming some of these challenges is the development and utilization of realistic autonomous power system breadboards and test beds on which KBES technologies can be developed and validated [3],[39]. Since space power systems involve new and highly dynamic technologies, it is through the development and subsequent use of testbeds that the necessary "engineering expertise" is being established and archived [18],[25]. Moreover, in order for autonomous power system development to proceed in a continuous manner researchers and developers must rely on the lower risk terrestrial testbeds as opposed to actual mission experience alone.

The primary autonomous power system functions that have been identified for application of knowledge-based systems include: status estimation, system health status monitoring and maintenance, fault detection and management, dynamic load scheduling and maintenance procedure advising [6]. Currently research and development is ongoing in almost all of these areas and proof of concept has been established by various prototype systems. One of the most crucial functions among these, since it is a first line defense against system or component failure, is the system health status monitoring. This is the latest area being researched for knowledge-based expert system applications.

During actual space missions, engineers must monitor the telemetry data from various power system sensor and identify and analyze any "significant findings." Significant findings are defined more deeply than those identified by most current systems; namely, the indication based on a single variable that a fault has occurred. Significant findings must be based on not only single variable values, but also on the interactions between the variables and the trends indicated by them. Such significant findings can indicate an imminent failure even though single variable analysis might not. Such a task will involve large amounts of data, only some of which (in many cases a very small percentage) will be relevant to any particular prediction. These factors, combined with the frequency and regularity of task

performance, requires engineering expertise on the functionality of the system, the mode of operation, and the efficient and effective reading of the telemetry data. Such a function can be most effectively implemented using knowledge-based systems technologies [29]. The application of such technologies will result in intelligent systems to support the monitoring function.

The main purpose of this paper is to review the literature applicable to the intelligent health status monitoring domain. The findings of this review will then be synthesized into recommendations for the development of intelligent space power monitoring systems. The review is broadened beyond space power system monitoring because many of the problem characteristics exist in other domains such as human health monitoring, manufacturing system monitoring, test data analysis and others. One characteristic of all such problems, however, is the requirement to efficiently and effectively perform what is referred to as "data reduction." Data reduction in this context is defined as the process of extracting from the larger amounts of monitoring data (usually being provided by real-time sensors) only the "significant" elements and presenting these elements to the analyst in a form most conducive to supporting decision making concerning the health status of a system. Reducing data to such a form usually involves the application of various statistical techniques such as graphing, plotting, calculating maximum and minimum values, calculating means, taking differences, determining trends, analyzing and comparing signatures, etc. The issue of intelligent data reduction has universal implications due to the amount of information now available to decision makers because of the advances in information processing and remote sensing technologies.

OBJECTIVES

The objective of this work is to investigate the application of knowledge-based system technologies to the field of space electrical power system health status monitoring. This objective was accomplished by first examining current applications for autonomous power systems with emphasis on the Hubble Space Telescope test bed and the Nickle Cadmium Battery Expert System - NICBES. Next a broad review of the literature related to knowledge-based monitoring systems employing intelligent data reduction techniques was conducted. In conjunction with the literature reviews, various NASA and contractor personnel were contacted or interviewed concerning the topic. Finally, the findings were synthesized into recommendations for future research efforts in this domain.

BACKGROUND

Marshall Space Flight Center's Electrical Power Branch has been involved since 1984 with the development of expert or knowledge-based systems [38]. Attention has been primarily focused on comprehensive fault management and dynamic payload rescheduling activities. Comprehensive fault management includes identifying anomalies, diagnosing actual faults, recommending corrective action for fault recovery and autonomous implementation of fault recovery actions. The knowledge-based systems which have been developed and are being researched as part of these efforts include: the Fault Isolation Expert System (FIES I and FIES II) [38], the Space Station Experiment Scheduler (SSES) [39], the fault detection/diagnosis/recovery system (STARR) [34], the Space Station Module Power Management and Distribution (SSM/PMAD) system automation project [6], the cooperative expert system project for Scheduling and Fault Analysis/Recovery Integration (SAFARI) [38], the Nickel Cadmium Battery Expert System (NICBES) [4],[21],[26] and the latest research efforts for Intelligent Data Reduction - I-DARE [17]. I-DARE is primarily being developed to enhance and extend the functioning of NICBES which currently interfaces with the Hubble Space Telescope (HST) power system testbed. Since the research presented in this paper was conducted to support and extend the efforts of the I-DARE project a brief overview of the three interfacing systems will be provided. For more detail on the other efforts listed the reader is referred to the cited references.

The HST (Hubble Space Telescope) Testbed was developed to simulate as close as practical the actual system that is to be flown on the HST [3] (see Figure 1). The testbed consists of six major elements: (1) the power distribution breadboard, (2) the batteries, (3) the solar array simulators (SAS), (4) the load banks, (5) the charge control hardware, and (6) a control computer (CC), a monitoring computer referred to as the Digital Data Acquisition System (DDAS), and a computer for the NICBES. The breadboard includes all pertinent components of the HST EPS. It provides switching for battery isolation, Solar Panel Array (SPA) switching and battery reconditioning, as well as monitoring via panel meters of individual battery voltage, current and temperature. Additionally, three strip chart recorders record voltage and current data.

Power storage is provided by six nickel cadmium batteries manufactured from the same lot as the actual flight hardware. During the sun portion of its orbit, the HST will be powered by 20 SPA's (three each per battery) which are simulated on the testbed by two adjustable constant current power supplies. Three independently controlled load banks simulate the spacecraft load and are controlled by the control computer. The charge control hardware consists of six

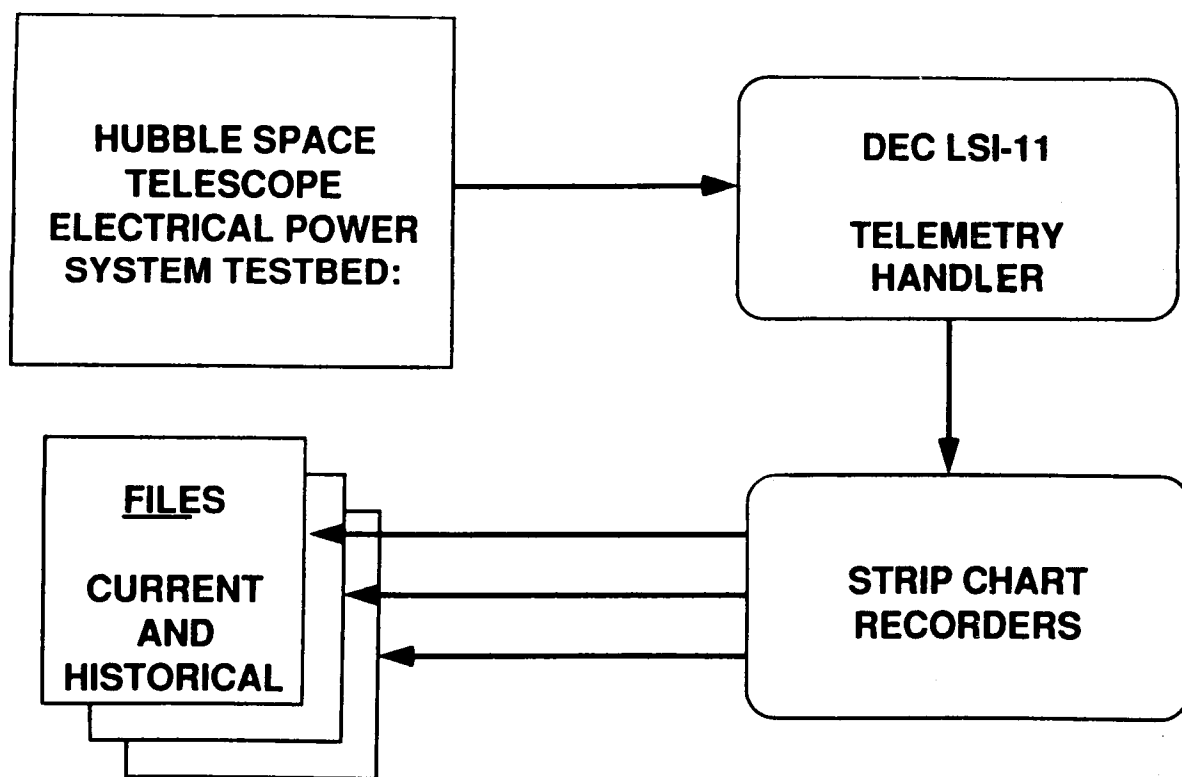


Figure 1 - HST Functional Diagram

ORIGINAL PAGE IS
OF POOR QUALITY

charge current controllers (CCC) identical in circuitry to the flight hardware. The Control Computer is a microprocessor based system designed and built at MSFC specifically for the breadboard. The Control Computer provides for keyboard entry, via a VDT, of several command adjustable parameters, thus providing functions such as: SAS control, CCC monitoring, and monitoring and control of load bank voltages, battery temperatures and battery voltages. The DDAS function is performed on a DEC LSI-11 computer and is responsible for data acquisition, limit checking, data summaries and orbit time control. The DDAS samples approximately 400 channels of incoming data from the breadboard. Once per minute it measures 138 cell voltages and pressures, currents for the six batteries and battery protection circuits, the three load voltages, six battery voltages and six temperatures per battery. The DDAS provides data summaries such as high/low reading, recharge ratio, depth of discharge, etc., on a per orbit basis. The third computer component of the testbed is currently as IBM PC/AT which is dedicated to the maintenance and functioning of the expert system, NICBES.

NICBES

The prototype of NICBES is currently integrated with the HST testbed and serves as a fault detection and diagnosis system and also a battery health management system [4] (see figure 2). Functionally NICBES has four modes: (1) fault diagnosis, (2) battery health status, (3) advice on battery maintenance, and (4) decision support aid. The prototype has two separate subsystems: a data handler and a diagnosis expert system. The data handler is written in Microsoft C and serves to receive the telemetry data from the DDAS and "massage" it into the form required by the diagnosis subsystem. The diagnosis subsystem is written in Arity PROLOG [26]. Its main function is to reason from the data provided by the data handler and determine if any exception situations are indicated. The engineer may initiate operation of the expert system once twelve orbits worth of data have been processed by the data handler. In addition to fault diagnosis which is based on the current state of the EPS, NICBES also monitors battery health status based on both current orbital data and historical data. Status analysis is accomplished based on the interactions of several variables. Based on an analysis of trends and averages for combinations of parameters maintenance procedures (i.e., battery reconditioning, charging scheme, etc.) are recommended. Finally, NICBES provides decision support to the user by supplying summary plots of pertinent data over the most current twelve simulated orbits.

The NICBES prototype has performed well over the past year and a half. During this period, however, several features have been identified for incorporation and upgrade of the system. Among these are: (1) the need for a user friendly rule editor, (2) a multitasking capability to allow data collection to continue while a NICBES consultation is in progress, (3) an expanded capability on the number of orbits of data which can be handled (upgrades from the current 12 to

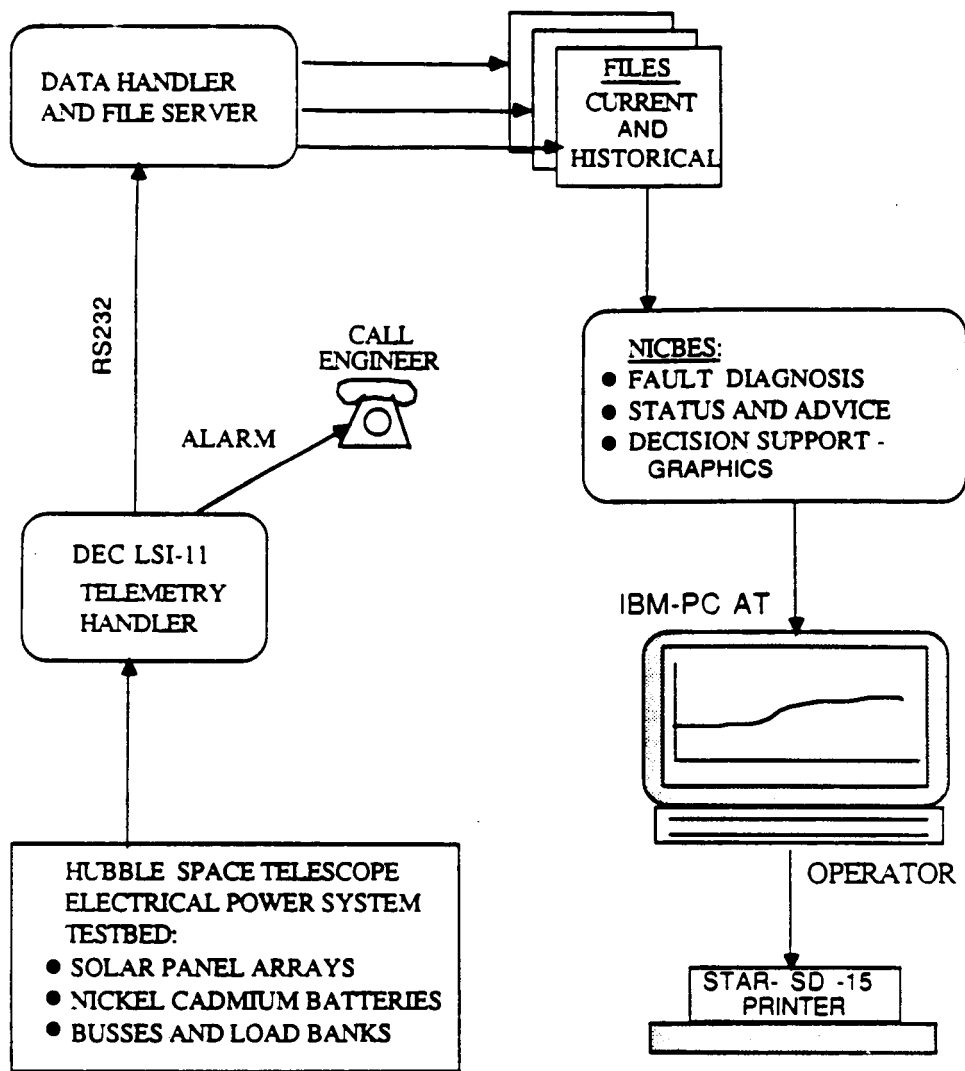


Figure 2 - NICBES Functional Diagram

perhaps 500 orbits have been discussed) (4) providing a capability to perform statistical prediction of battery life, and (5) the growth of NICBES into a model based system to explicitly capture the interrelationships of the EPS and thus allow diagnosis of unforeseen faults. All of these features are currently being researched and evaluated.

I-DARE

Data reduction takes many forms, the most common form is statistical reduction, i.e., means, moving averages, trend analysis, minimums and maximums, etc [1],[14],[32]. Though these are very useful for analyzing system performance further reduction which considers relationships between variables as well as system functioning lends support to an even deeper analysis. Such reduction brings more of a qualitative approach to the problem. This can be referred to as "intelligent" reduction. Intelligent because expertise is brought to bear on the problem through the reduction process and also because the data is refined by considering the interactions of the system being analyzed. The overall purpose of the I-DARE effort is to investigate the phenomenon of data reduction, determine the knowledge used in performing this task by humans and prototyping a system for incorporation into the HST testbed [17].

Currently the amount of data being received from the DEC LSI-11 included 370 values per minute. Since each orbit is ninety-six minutes in duration the orbital analysis must be performed on 35,520 values. Adding to the amount of data is the fact that 100 orbits of data are usually required for accurate trend analysis. Such analysis, therefore, must be based on 3,552,000 values. Complete analysis of such a quantity of data is a formidable task for anyone and is complicated even further when all the interrelationships are considered. The formidability of the monitoring task coupled with large demands on computer storage require advanced forms of intelligent data reduction to support engineers in reducing power system telemetry data to its significant components. The next section will provide a review and summary of the findings of current efforts in applying knowledge-based technologies to the health status monitoring function.

CURRENT KNOWLEDGE-BASED APPROACHES TO SYSTEM HEALTH STATUS MONITORING

Over the last three to four years several efforts have been undertaken to investigate and prototype knowledge-based systems for the health and status monitoring of space systems. This section will briefly review these efforts and summarize the findings. A thorough understanding of current system monitoring technologies and their approaches to data reduction is a necessary prerequisite to the design and implementation of effective data reduction techniques for further improving the monitoring function.

Siemens [30] has written several papers on a knowledge-based

system, STARPLAN, which monitors telemetry data, alerts the operator if anomolous conditions are identified, and then functions to suggest corrective actions. STARPLAN was built using the KEE environment. KEE facilitated development of the knowledge base since the knowledge base consists of object descriptions which are reasoned about based on satellite telemetry. It performs fault diagnosis by utilizing relational links between objects of the domain model and the descriptions of the objects. It uses production rules only if the model-based approach fails.

Hamilton [22] implemented a system in LISP on a Symbolics 31670 which was called SCARES and was applied to the attitude control system of a spacecraft. SCARES was implemented with an object oriented approach using frame-based knowledge representations and inductive and deductive reasoning. It uses a three stage approach of monitoring, diagnosis and hypothesis generation and test to detect, diagnose and recover from anomalies. The monitoring is done by performing three types of checks on the telemetry: a limit check on individual telemetry points, a rate check on two or more telemetry paint on the same channel, and a cross-channel check for consistency. Rather than analyzing all the data, the monitor only receives a sample of each signal once every two seconds. These samples are sufficient to monitor and detect faults with real time performance.

Skapura and Zach [31] describe an OPS5 based system developed using LISP and designed to handle the front-end analysis of the telemetry stream received from the Space Shuttle. As a result of their investigation into application of knowledge-based systems for real-time analysis, they draw several important conclusions. They point out the problems encountered in attempting to use OPS5 for such applications. Specifically performance limitations, lack of interrupt handling facilities and the need for a multi-tasking architecture. The authors also point out as do Watson, Russell and Hackler [] that the RETE algorithm, designed to work in environments where the data changes slowly, is not optimal for real-time telemetry data analysis.

Gholdston Janik and Lane [20] report on a prototype expert system to aid in the evaluation of sensor data to monitor and predict power component performance and to identify faults. The system was developed on a Compag 80386 computer (with a 80387 co-processor) using the M1 rule-based development tool. As well as providing a rule-based language, M1 was also chosen for its ability to interface with the data acquisition, reduction and graphics routines which were done in C. The system is designed to operate in an automatic monitoring mode or an interactive diagnostic mode. During automatic monitoring mode, all EPS information is collected from the distributed processors and made available to the expert system through a data base. The expert system then determines if any of the data are exhibiting a failure mode.

Pooley, Thompson, Hamsley and Teoh [28] discuss the architecture of an intelligent health monitoring system (HMS) for reusable rocket

engine systems. SEES, SPARTA Embedded Expert System, synergistically integrates vibration analysis, pattern recognition and communications theory techniques with AI techniques. The main component of the application is an expert system that uses confidence levels to resolve conflicts among compound data and then heuristically chooses each data set and derives classification rules. The system is comprised of three major subsystems. The SEES Front End (SFE) processes raw data to screen obvious anomalies and derive the reduced data set from which it generates an appropriate signature (the authors do not detail the data screening, reduction and signature generation techniques since this information is considered propriety). The Embedded Expert System (EES) uses the information provided by the SFE and the rule set in its knowledge base to infer operating conditions, deduce mean time to failure and recommend maintenance schedules. The EES has the capability to invoke functions in the SFL for further data reduction. The expert system component is being developed using Rule Master which can then be integrated with the rest of the system which is developed in C. The Support Function Library (SFL) is a set of supporting functions for the rest of the HMS.

Watson, Russell and Hackler [36] report on the design and initial development efforts of the Diagnosis and Protection Expert System (DAPES), an expert system for the purpose of performing on-line diagnostics and parameter evaluation to determine potential or incipient fault conditions in electrical power systems. The system will be part of an overall monitoring computer hierarchy to provide a full evaluation of the status of the power system and react to both incipient and catastrophic faults. The attempt with DAPES is to provide as much computational intelligence as possible to the remote low-level machinery as possible. Such a capability is feasible with current advances in microprocessor hardware technologies. The authors' efforts are focused on the architecture of a responsive expert system for on-line monitoring environments.

To accomplish this responsiveness an expert system shell PMCLIPS was developed. PMCLIPS is a modification of CLIPS. It uses a Parallelized Match Algorithm (PMA) based on the RETE Algorithm. Though some of the internal data holding structures are similar for the two algorithms, the flow through the structures varies considerably. PMCLIPS takes advantage of the rapidly changing data and parallelism which characterize monitoring systems to enhance the system's responsiveness. Initial results indicate that significant speed up is accomplished. Prototyping and testing is continuing. Though the authors site data compression and detection of incipient faults as functions of DAPES, no detail is provided concerning these functions.

Doyle, Sellers and Atkinson [11] present an approach to monitoring referred to as "Predictive Monitoring" which is based on the idea that effective monitoring requires an explicit model of a device. The model requirement is based on (1) the fact that the nominal ranges associated with a sensor are dynamic as opposed to static as in traditional

monitoring approaches, and (2) the fact that as space systems become more and more complex and sensors number in the thousands, they must be treated as resources and context-sensitive importance criteria should be used to determine when and how they should be sampled. Additionally, the authors feel that providing explanation for expected sensor values would speed up the diagnosis process.

A predictive modeling system called PREMON is proposed and is made up of three modules: (1) a Causal Simulator, (2) a Sensor Planner, and (3) a Sensor Interpreter. The causal simulator, based on a model of the system being monitored, generates predictions about the next cycle of behavior of the device. The model distinguishes different operating modes of the device and compliments traditional analytical models with qualitative reasoning capabilities to accommodate causal dependencies and incomplete and uncertain values. The Sensor Planner, given the predicted behavior, makes choices about what behavior to verify, which sensors to employ, and how the sensors should be sampled. The Sensor Planner then passes instructions to the Sensor Interpreter which reads sensor channels as directed and compares these with the expectations provided by the Causal Simulator. Expectations about behavior and knowledge about distinguishable qualitative values, which are derived from a device model, drive the comparison and recognition process.

Erickson and Rudakas [15] also report on the advantages to be gained from a model-based reasoning approach for knowledge-based application to monitoring and fault diagnosis systems. The paper discusses the two-phase development of a system called TEXSYS, Thermal Expert System, which was developed around the Space Station Thermal Testbed, SSTT (again illustrating the major importance of testbeds to the development of KBS components for autonomous systems). TEXSYS was first prototyped using KEE and SimKit on a Symbolics 3670. The phase II prototype employed MTK, the Ames Research Center's causal Model Toolkit, in place of SimKit. MTK was designated to overcome several technical issues raised during phase 1 development. Several key features of the TEXSYS system are worthy of mention.

The SSTT model which was developed using the object-oriented and frame-based representation of KEE was a simplified model of the testbed which incorporated domain experts' rules of thumb, as well as the relevant physical laws. Additionally, rather than using direct links, the model topology was represented using KEE CONNECTIONS which allowed representation of behavior at component boundaries. KEE worlds was used to represent different temporal states in distinct "worlds" thus providing a limited temporal reasoning facility. A hierarchical structure was employed to allow reasoning to progress from "black box" levels to subcomponent level. The ability to perform qualitative and quantitative modeling and reasoning, as well as the ability to deal with parameter uncertainty was provided by the structure of MTK. MTK allows for utilization of "Parameters" which represent the significant physical measurements employed in describing a given system (i.e., temperature, pressure, flow rate, etc.). Parameters are represented as

special objects with Value, Best Value, State and History attributes. Values can be given or derived quantitative measures or ranges and can take on multiple values reflecting different chains of inference. Best Value represents the systems best guess of the "real value." The State is simply a qualitative symbolic value of a Parameter (i.e., NOMINAL, HIGH, NEGATIVE, etc.) and History, another symbolic value, captures time behavior such as INCREASING, STEADY or DECREASING.

The most current and perhaps most significant research efforts are those reported by various personnel at the Lockheed Artificial Intelligence Center [7], [12], [13]. Lockheed is the prime contractor for the Hubble Space Telescope, moreover, it is the HST Mission Operations Contractor responsible for the ground operations that ensure the health and safety of the vehicle. The primary means for assuring such operation is the effective monitoring of approximately 4,690 different telemetry monitors available for interpretation. Six workstations will be manned 24 hours a day by three shifts of operators during actual flight. The complexity of the HST and the massive amounts of telemetry data being received make system health status monitoring extremely difficult. Lockheed, therefore, has been actively engaged in applying KBS technologies to this domain. The main thrust of these efforts are directed toward development of the Telemetry Analysis Logic for Operating Spacecraft (TALOS) system.

TALOS is a knowledge-based system consisting of a multitasking architecture for performing real-time monitoring and off-line deep analysis. It is being developed using Lockheeds L*STAR proprietary shell. The system consists of three separate processes which run concurrently and communicate via message passing using a mailbox approach. The three processes are the Inference Process, the Data Management Process (DMP) and the I/O Process.

Telemetry data is first preprocessed by a VAX and is then sent to the DMP for scaling and compression before utilization by the Inferencing and I/O Processes. The Inference Process is written in C and has been designed to overcome certain deficiencies of other tools when applied to the monitoring domain. In order to reduce long run-time pattern matching searches, the rule compiler stores location information about triples in a uniform data structure created at compile time. There is also a context mechanism (context relating to mode of HST operation for example) which partitions rules to limit the number being examined for given situations. This context sensitivity allows "attention" to be focused when important events occur (i.e., the context may cause a rule to fire which increases the rate at which the DMP sends data on a particular monitor). The Inference Process uses special functions which reason about trends and statistics such as time averages and rates-of-change to provide temporal reasoning capabilities about past, present and future events. The Inference Process also contains the "knowledge" required by the DMP and sends this intelligence as necessary via messages. The DMP needs to know for each telemetry monitor such things as sampling rate smoothing and scaling factors,

limits, etc. These factors are dynamic based on the requirements as reflected in the knowledge base.

Also resident in the Inference Process are the diagnostic modules which are activated periodically to analyze archived telemetry data or automatically when non-nominal operation is indicated based on real-time telemetry analysis. Archived data is analyzed for anomalous behavior as well as adverse trends. For real-time analysis only about 400 of the over 4000 telemetry monitors are used by operators under normal operation. If non-nominal behavior is indicated, then rules fire in the knowledge base resulting in messages being sent to the DMP to request data collection from new monitors and/or changes in sampling rates or data compression techniques. The I/O Process is a hierarchy of displays which can be traversed with a mouse. Mousing on a monitor displays a strip chart which is updated in real time by data from the DMP. The I/O Process also receives status/health messages from the Inference Process.

Findings

A synthesis of the findings from the literature reviewed above leads to several important observations and conclusions concerning the effective application of knowledge-based technologies to system health status monitoring. Those which have relevance to the concept of data reduction and which should be considered in the implementation of reduction techniques in this domain are now reviewed.

New technologies and complex systems overburden analysts with telemetry data and make manual real-time or near real-time analysis infeasible. Automating portions of the monitoring process has proven feasible but traditional approaches of establishing a-priori static nominal ranges proves ineffective for most current applications where the changing contexts of the system dictate the nature of analysis and the data required to support this analyses. Sensors must be treated as information resources and managed accordingly. The context should dictate the what, when and how to monitor. One approach to accomplishing this is with a model-based system where an explicit system model which captures the causal relationships and dependencies is part of the knowledge base. The model expectations can then help drive the analysis.

Effective monitoring systems will have to possess reasoning capabilities beyond just quantitative reasoning. More specifically, qualitative and temporal reasoning are considered mandatory. Qualitative reasoning can accommodate uncertain missing and noisy data [19]. Additionally, the temporal relationships between system components and the data generated must be captured by the system. Trend detection and analysis are crucial to effective system health status monitoring.

Finally, the rapidly changing data inherent with space system

monitoring dictates certain system characteristics. Systems must have interrupt or multitasking capabilities. Moreover, the inferencing process must be freed of data management functions and I/O functions. This can be accomplished with a system architecture which provides for three separate modules which accomplish these tasks and communicate with each other as necessary. The rapidly changing data also requires new and innovative inferencing techniques to overcome the slow search and matching algorithms characteristic of most current KBS tools.

Implications for Intelligent Data Reduction

As stated previously, intelligent data reduction is the process of extracting from a set of data only those significant facts which the "expert(s)" deem necessary for the analysis being undertaken. The concern in this investigation was to determine the types of characteristics that might be deemed significant for the health status monitoring of space power systems and potential data reduction techniques capturing these characteristics. In considering data reduction techniques to support health status monitoring, one must also consider the operation of fault detection and diagnosis which often follows the monitoring function. One major distinction between these functions is the time frame for performance. Monitoring is something that must be done on a near real-time basis so that anomalies can be detected, or, better yet, predicted early enough to prevent further system contamination. It has been clearly shown that the efficiency of the monitoring process can be greatly improved by first performing data reduction on the telemetry data [5] [24] [28] [33].

Moreover, much of the information which would be considered significant for health status monitoring can also play a major role in anomaly identification. It has been shown that the most effective approaches to fault detection involve model-based expert systems. This same model-based system can be used to drive a data reduction module.

Causal modeling allows the interrelationships and dependencies of system components to be captured from both a physical system perspective (i.e., based on physical laws), as well as a conceptual perspective (based on the intuitive and heuristic knowledge of an expert) [2] [15] [35]. Such an approach also allows system contexts to be considered since the model can capture the various defined states in which the system will operate. Often as contexts change, the interrelationships of the model will change and vice-versa.

Data reduction techniques should reflect these relationships. For example, during battery charging, cell temperature is a telemetry value which needs to be closely monitored. Consequently, during a charge state, the temperature sensor data should be sampled more frequently than during a discharge state. Likewise, cell pressure which is related to temperature should also be monitored at a similar rate. Another advantage of the modeling approach is that it can provide a means of checking for telemetry data contamination or loss. The

relationships which exist in the model can be used to predict or verify the various telemetry values. In terms of system structure for such an approach, a separation of reduction and inferencing functions is most appropriate. A library of reduction functions as specified by the analyst might be incorporated into the data handler and could receive and pass necessary information from and to the model-based inferencing module.

The hierarchical approach mentioned previously also has applications to the data reduction function. It would seem natural that as the monitoring function proceeds from broader (i.e., a conceptual black box level) to finer (i.e., subsystem or component level), levels of detail, the data reduction employed will also go to finer levels of detail (i.e., from a qualitative measure such as INCREASING to a more detailed and quantified trend analysis) [19]. At the conceptual level, overall battery readings might be monitored, whereas at a lower level, individual cell values would be of concern. The level of detail would again be influenced by the context as well as the model interactions.

CONCLUSIONS AND RECOMMENDATIONS

This research has clearly shown that the efficiency of the health status monitoring function can be greatly improved using data reduction techniques on the telemetry data. It has been pointed out, however, that due to the complexity of current space power systems and the dynamics of the environments in which they operate, monitoring systems can no longer be static in nature. It appears that intelligent health status monitoring systems must go beyond complete rule-based systems to hierarchically structured model-based systems. Such an approach allows the inherent dynamics of the system itself, as well as the dynamics of the system in different contexts or modes, to be reflected in the monitoring system and the supporting data reduction subsystem or module.

It has also been shown that much of the reduction process should take the form of converting quantitative telemetry data to qualitative data. Knowledge-Based Systems gain their advantage because they can perform symbolic reasoning. Thus the best approach to data reduction for such applications is to process the data using traditional computing environments with the goal of reducing the data to qualitative symbolic representations which can then be reasoned about in the knowledge base. Merging qualitative and quantitative data analysis in knowledge-based systems for EPS health status monitoring is an area for further investigation.

REFERENCES

1. Berington, Philip R., Data Reduction and Error Analysis for the Physical Sciences, McGraw Hill, New York, 1969.
2. Blasdel, Arthur N., Jr., "Automated Fault Handling of a Satellite Electrical Power Subsystem Using a Model-Based Expert System", Proceedings of the 22nd IECEC, Philadelphia, PA, 1987, pp. 601-606.
3. Bush, John R., Lorna G. Jackson and John R. Lanier, Jr., "Hubble Space Telescope Electrical Power System Simulation Breadboard", Proceedings of the 22nd IECEC, Philadelphia, PA, 1987, pp. 618-622.
4. _____, "Final Report for Nickel Cadmium Battery Expert System", Report No. MCR-85-641, Martin Marietta Aerospace Division, Denver, CO, November 1986.
5. Cralene, Robert, "Systems and Methods to Reduce Data Processing Turnaround Time", Proceedings of the International Telemetry Conference, Los Vegas, NV, October 1986, pp. 391-396.
6. _____, "Task 1. Study Reports on Space Station Automation of Common Module Power Management and Distribution", Report No. MCR-86-583, Martin Marietta Aerospace Division, Denver, CO, July 1986.
7. Cruse, Bryant G., "TALOS: A Distributed Architecture for Intelligent Monitoring and Anomaly Diagnosis of the Hubble Space Telescope", Proceedings of the Third Conference on Artificial Intelligence for Space Applications, Huntsville, AL, 1987, pp. 103-107.
8. Decker, D. K., "A Methodology for Designing Fault Tolerant Spacecraft Subsystems", Proceedings of the 22nd IECEC, Philadelphia, PA, 1987, pp. 672-678.
9. Dolce, James L. and Karl A. Faymon, "Automating the U.S. Space Stations Electrical Power System", Optical Engineering, November, 1986, Vol. 25 No. 11, pp. 1181-1185.
10. Donovan, R.M. and L. Song, "An Expert System for Control and Data Reduction in Cytometry", Proceedings of the Ninth Annual IEEE Engineering in Medicine and Biology Conference, Boston, MA, November 1987, pp.

1553-1554.

11. Doyle, Richard J., Suzanne N. Sellers and David J. Atkinson, "Predictive Monitoring Based on Causal Simulation", Proceedings of the Second Annual Artificial Intelligence Research Forum, Palo Alto, CA, 1987, pp. 44-59.
12. Dunham, Larry L., et al., "Knowledge-Based Monitoring of the Painting Control System on the Hubble Space Telescope", Proceedings of the Third Conference on Artificial Intelligence for Space Applications, Huntsville, AL, 1987, pp. 103-107.
13. Eddy, Pat, "A Knowledge-Based System for Monitoring the Electrical Power System of the Hubble Space Telescope", Proceedings of the Third Conference on Artificial Intelligence for Space Applications, Huntsville, AL, 1987, pp. 103-107.
14. Ehrenberg, A.S.C., Data Reduction - Analysing and Interpreting Statistical Data, John Wiley, New York, 1975.
15. Erickson, William K. and Mary R. Rudokas, "MTK: An AI Tool for Model-Based Reasoning", Proceedings of the 2nd Annual Artificial Intelligence Research Forum, Palo Alto, CA, 1987, pp. 130-134.
16. Faymon, Karl A., Gale R. Sundberg, Robert R. Bercaw and David J. Weeks, "LERC Power System Autonomy Program - 1990 Demonstration", Proceedings of the 22nd IECEC, Philadelphia, PA 1987, pp. 547-551.
17. Ford, Donnie R. and David J. Weeks, "Intelligent Data Reduction: A Preliminary Investigation", Proceedings of the 23rd IECEC, Denver, CO, 1988, pp. 383-388.
18. Freeman, Michael S., "HSTDEK: Developing a Methodology for Construction of Large-Scale, Multi-Use Knowledge Bases", Proceedings of the Third Conference on Artificial Intelligence for Space Applications, Huntsville, AL, 1987, pp. 89-94.
19. Ganascia, J.G., "Using an Expert System in Merging Qualitative and Quantitative Data Analysis", International Journal of Man Machine Studies, Vol. 20 No. 3, March 1984, pp. 319-330.
20. Gholdston, Edward W., Don F. Janik and Garth Lane, "A Diagnostic Expert System for Space-Based Electrical

- Power Networks", Proceedings of the 23rd IECEC, Denver, CO, 1988, pp. 401-406.
21. Glass, Betty, "Prototype for the Automation of Electrical Power Systems", Proceedings of the 22nd IECEC, Philadelphia, PA, 1987, pp. 552-556.
 22. Hamilton, Marc, "SCARES - A Spacecraft Control Anomaly Resolution Expert System", In N.K. Karnel, et al., Editors, Expert Systems in Government Symposium, 1986, pp. 436-443.
 23. Heer, Ewald and Henry Lum, "Raising the AIQ of the Space Station", Aerospace America, 1987, pp. 16-17.
 24. Kao, S., et al., "Real-Time Analysis of Telemetry Data", in N.K. Karnel, et al., editors, Expert Systems in Government Symposium, 1987, pp. 137-144.
 25. Keller, Richard M., Edward A. Figenbaum and Bruce Buchanan, "Development of a Reusable Knowledge Base for Space Applications", Proceedings of the Second Annual Artificial Intelligence Research Forum, Palo Alto, CA, 1987, pp. 357-364.
 26. Kirkwood, Nancy and David J. Weeks, "Diagnosing Battery Behavior with an Expert System in Prolog", Proceedings of the 21st IECEC, San Diego, 1986, pp. 1801-1807.
 27. Paz, Naemi, Cloyd Ezell and Dia Ali, "Heuristics for a Robot Data Explorer", Computers and Industrial Engineering, VII, 1986, pp. 1-4.
 28. Pooley, J., et al., "Embedded Expert System for Space Shuttle Main Engine Maintenance", Proceedings of the Third Conference on Artificial Intelligence for Space Applications, Huntsville, AL, 1987, pp. 115-119.
 29. Prerau, David S., "Selection of an Appropriate Domain for an Expert System", The AI Magazine, Summer 1985, pp. 26-30.
 30. Siemans, R. W., Marilyn Golden and J. C. Ferguson, "StarPlan II: Evolution of an Expert System", Proceedings of AAAI-86, 1986.
 31. Skapura, David M. and David R. Zoch, "A Real-Time Production System for Telemetry Analysis", in N.K. Karnel, et al., editors, Expert Systems in Government Symposium, 1986, pp. 203-209.

32. Thisted, Ronald A., "Representing Statistical Knowledge for Expert Data Analysis Systems", in William A. Gale ed., Artificial Intelligence in Statistics, Addison-Wesley, Reading, MA, 1986.
33. Utt, W.K., et al., "An Expert System for Data Reduction", Proceedings of the Second Conference on Artificial Intelligence Applications: The Engineering of Knowledge-Based Systems, Miami Beach, FL, 1985, pp. 120-124.
34. Walls, Bryan, "Starr: An Expert System for Failure Diagnosis in a Space Based Power System", Proceedings of the 23rd IECEC, Denver, CO, 1988, pp. 303-306.
35. Walters, John and Norman R. Nielsen, Crafting Knowledge - Based System, John Wiley and Sons, NY, 1988, pp. 285-301.
36. Watson, Karan, Don Russell and Irene Hackler, "Expert System Structures for Fault Detection in Spaceborne Power Stations", Proceedings of the 23rd IECEC, Denver, CO, 1988, pp. 389-394.
37. Weeks, David J., "Artificial Intelligence and Space Power Systems Automation", Proceedings of the Third Conference on Artificial for Space Applications, Huntsville, AL, 1987, pp. 109-113.
38. Weeks, David J., "Artificial Intelligence Approaches in Space Power Systems Automation at Marshall Space Flight Center", Proceedings of the First International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Tullahoma, TN, 1988, pp. 361-366.
39. Weeks, David J., "Space Power System Automation Approaches at the George C. Marshall Space Flight Center", Proceedings of the 22nd IECEC, Philadelphia, PA, 1987, pp. 538-543.

N89-21738

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

EXPERIMENTAL RESULTS ON ATOMIC OXYGEN CORROSION OF SILVER

Prepared by:	Albert T. Fromhold, Ph.D.
Academic Rank:	Professor
University and Department:	Auburn University Department of Physics
NASA/MSFC:	
Laboratory:	Materials and Processes
Division:	Engineering Physics
Branch:	Physical Sciences
MSFC Colleague:	Ann F. Whitaker
Date:	August 15, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

EXPERIMENTAL RESULTS ON ATOMIC OXYGEN CORROSION OF SILVER

by

Albert T. Fromhold
Professor of Physics
Auburn University
Auburn University, AL 36849

ABSTRACT

The results of an experimental study of the reaction kinetics of silver with atomic oxygen in 10° increments over the temperature range 0° to 70°C is reported. The silver specimens, of the order of $10,000\text{ \AA}$ in thickness, were prepared by thermal evaporation onto 3-inch diameter polished silicon wafers. These were later sliced into pieces having surface areas of the order of $1/4$ to $1/2$ square inch. Atomic oxygen was generated by a gas discharge in a commercial plasmod asher operating in the megahertz frequency range. The sample temperature within the chamber was controlled by means of a thermoelectric unit. Exposure of the silver specimens to atomic oxygen was incremental, with oxide film thickness measurements being carried out between exposures by means of an automated ellipsometer. For the early growth phase, the data can be described satisfactorily by a logarithmic growth law: The oxide film thickness increases as the logarithm of the exposure time. Furthermore, the oxidation process is thermally activated, the rate increasing with increasing temperature. However, the empirical activation energy parameter deduced from Arrhenius plots is quite low, being of the order of 0.1 eV .

ACKNOWLEDGEMENTS

Acknowledgements are owed to many people, among them the following:

- * **Mrs. Ann Whitaker**, MSFC Colleague, for helpful discussions and for use of laboratory equipment, especially the plasma asher and thermoelectric unit
- * **Ralph Carruth**, Branch Chief, for his interest and useful suggestions
- * **John Reynolds**, for aiding me in establishing communication with the automated ellipsometer
- * **Bob DeHaye**, for helpful advice and stimulating ideas
- * **David Esker**, Ga. Tech. co-op student, who often lent a hand with the measurements
- * **Many other members of the Engineering Physics Division** under the direction of Dr. Ray Gause, for help, advice, and good fellowship
- * **Bob Schwinghamer**, Materials and Processes Laboratory Director, for his interest, encouragement, and insightful comments
- * **Ron Beshears**, Auburn graduate student, for preparing the evaporated silver specimens
- * **Dr. John Williams**, Auburn University Professor, for taking RBS spectra of the oxidized silver specimens
- * **Dr. Mike Freeman**, University Program Co-Director, for his generous organizational style and good humor
- * **Mrs. Ernestine Cothran**, MSFC Program Co-Director, for her sincere interest and careful attention to detail

I. INTRODUCTION

There is very little atomic oxygen in the ambient atmosphere on earth. That which is produced by decomposition of molecular oxygen by incident ultraviolet radiation from the sun which penetrates the earth's atmosphere undergoes rapid recombination to molecular oxygen because of the high collision rates at atmospheric pressure. The situation is quite different at very high altitudes, where the ultraviolet flux is relatively greater and the pressure is relatively much lower. The molecular oxygen is then decomposed much more rapidly, and the recombination rate due to atomic collision is vastly reduced. The primary form for oxygen at altitudes of low earth orbit trajectories is therefore atomic. One might initially think that the low ambient pressure at low earth orbit trajectories would imply little gas-solid interaction, but this is not true. The interaction of atomic oxygen with various materials has important implications for NASA's space telescope and space station programs.

The basic problem is that atomic oxygen is so much more reactive than molecular oxygen that corrosion effects on materials are greatly worsened in low-earth orbit, relative to corresponding rates on earth. One of the best examples of this is the deterioration of the elemental metal silver while exposed on board the space shuttle (1,2). Silver is swiftly attacked by atomic oxygen to form a black oxide layer which subsequently cracks and spalls. The purpose of the presently-described experimental work was to study the very early stages of the oxidation of silver by atomic oxygen.

The Physical Sciences Branch of the Engineering Physics Division of the Materials and Processes Laboratory of Marshall Space Flight Center has a renowned history in studying atomic oxygen effects on various materials, including elemental metals, alloys, organics, epoxy resins, paints, etc. The laboratory personnel have carried out flight experiments, supplemented by many ground-based experimental studies (1,2). The measurements described herein were carried out in collaboration with the members of that laboratory using their facilities.

II. OBJECTIVES

The primary objective of this study was to determine experimentally the early-stage kinetics of the reaction of atomic oxygen with silver. This requires determination of the oxide film thickness as a function of exposure time at a given temperature, with a series of temperatures employed to determine the temperature dependence of the kinetics. The secondary objective was to explore the theoretical implications of the experimental results.

III. EXPERIMENTAL

A. Procedures

Polished silicon wafers were used as substrates for evaporated silver layers 10,000 Å in thickness, as measured by a Dektak mechanical profiler. After evaporation, the wafers were cut into smaller specimens 1/4 to 1/2 inch on edge. Before each temperature run, the optical constants of the silver specimen were measured with a Gaertner automated ellipsometer having a laser ($\lambda = 6328 \text{ Å}$) light source for producing the measuring beam. Typical values for the refractive index and absorption coefficient were $n_s = 0.1$, $k_s = 3.8$. Reaction with atomic oxygen produced a surface oxide having typical optical constant values of $n_f = 1.8$, $k_f = 0.1$, though a range of values [$(1.6 \leq n_f \leq 2.2)$; $(0.05 \leq k_f \leq 0.25)$] for each was required to satisfy the minimum error criterion for different thicknesses.

Specimens were placed in the chamber of a commercial plasma asher which was pretuned at the operating conditions of 60 Watts and 800 millitorr oxygen pressure. The flux of atomic oxygen under these conditions has been estimated (3) to be $10^{18}/\text{cm}^2\text{-sec}$. After locating the specimen on the thermoelectric temperature controller in the chamber, evacuation was carried out by a vacuum pump operating with a nonorganic fluid, and the temperature of the specimen was regulated by adjusting the voltage on the thermoelectric device. A constant-temperature chilled solution was circulated continuously through the mounting block for the thermoelectric device and specimen in order to provide the reference temperature and a heat source and sink. The sample temperature was monitored with a thermocouple. Oxygen pressure was varied by adjusting the rate of steady-state flow of molecular oxygen into the chamber. After achieving the desired temperature, the oxygen pressure was raised and stabilized at 800 millitorr, and the gas discharge initiated. Initial specimen exposures were in half-second increments, but the length of the exposure increment was steadily increased as the reaction rate decreased. Initial exposure periods were required to be this short because of the very high initial reaction rates. However, there is definitely some uncertainty in the timing of such extremely short bursts of the gaseous discharge. After each exposure, the specimen was allowed to return to ambient temperature in the reaction chamber. It was then removed from the chamber, placed in air on the stage of the automated ellipsometer, and the surface oxide film thickness measured. For reasons which become apparent in the next section, especial care was taken to shutter the laser light to prevent exposure of the sample except during the short time increment (20 seconds or so) required for the ellipsometer to collect the optical data needed for computation of the oxide thickness in a given measurement. Following determination of the thickness, the sample was returned to the plasmod reaction chamber, and the cycle repeated.

B. Interaction of Laser Beam and the Oxide Film

An unexpected interaction was found between the laser light beam of the ellipsometer used for oxide film thickness measurement and the oxide film itself, namely, the readings of the ellipsometer for the film thickness were found to change slowly but continuously if the light beam was reflected continuously from the sample during and between individual measurements. This effect is shown in the bar graph for oxide film thickness in Fig. 1, where it can be noted that the thickness reading continuously decreases with the measurement number while the light is continuously reflected from spot A on the specimen. Moving the sample stage so that the light is reflected from a new position (spot B) yields initially a higher thickness reading, but that reading then also continuously decreases with the measurement number.

Such an interaction between observer (measuring instrument) and the object (that which is observed) is expected on a microscopic level, being the *raison d'être* of the Heisenberg uncertainty relation. However, it was surprising to us to find the effect on such a macroscopic level. This is perhaps indicative of the relative instability of silver oxide, especially under the uncontrolled ambient conditions in air while the sample was on the ellipsometer stage during measurement. It could even be associated with changing optical properties of the oxide because of continuous exposure to the light beam. Irrespective of the nature of the observed effect, our practical solution to this problem was quite straightforward -- we simply shuttered the light beam from the sample being measured except for the brief period required for the ellipsometer to collect the data in each measurement.

C. Oxidation Kinetics

Figure 2 illustrates the increase in oxide film thickness with time of exposure to atomic oxygen for a silver specimen held at a temperature of 20°C during exposure. Note from this figure that the oxide thickness initially increases very rapidly with exposure time, but this is followed by a leveling off to a slow growth region. The observed behavior was found to be typical of the data taken at all temperatures utilized in this study, namely, 0, 10, 20, 30, 40, 50, 60, and 70°C. Not all data, however, were taken over such an extended exposure period as that for the data in Fig. 2, which can be noted to extend out to a cumulative value of 280 seconds. The maximum oxide film thickness obtained for this exposure can be noted to be of the order of 2600 Å.

Corresponding data taken at a temperature of 40°C are shown in Fig. 3. This curve exhibits also the initial rapid growth phase, followed by the leveling off to a much slower growth period. The maximum exposure time for the 40°C data can be noted to be 26 seconds, at which time the oxide film thickness can be noted to be approximately 2000 Å.

Laser Photodecomposition - Silver Oxide

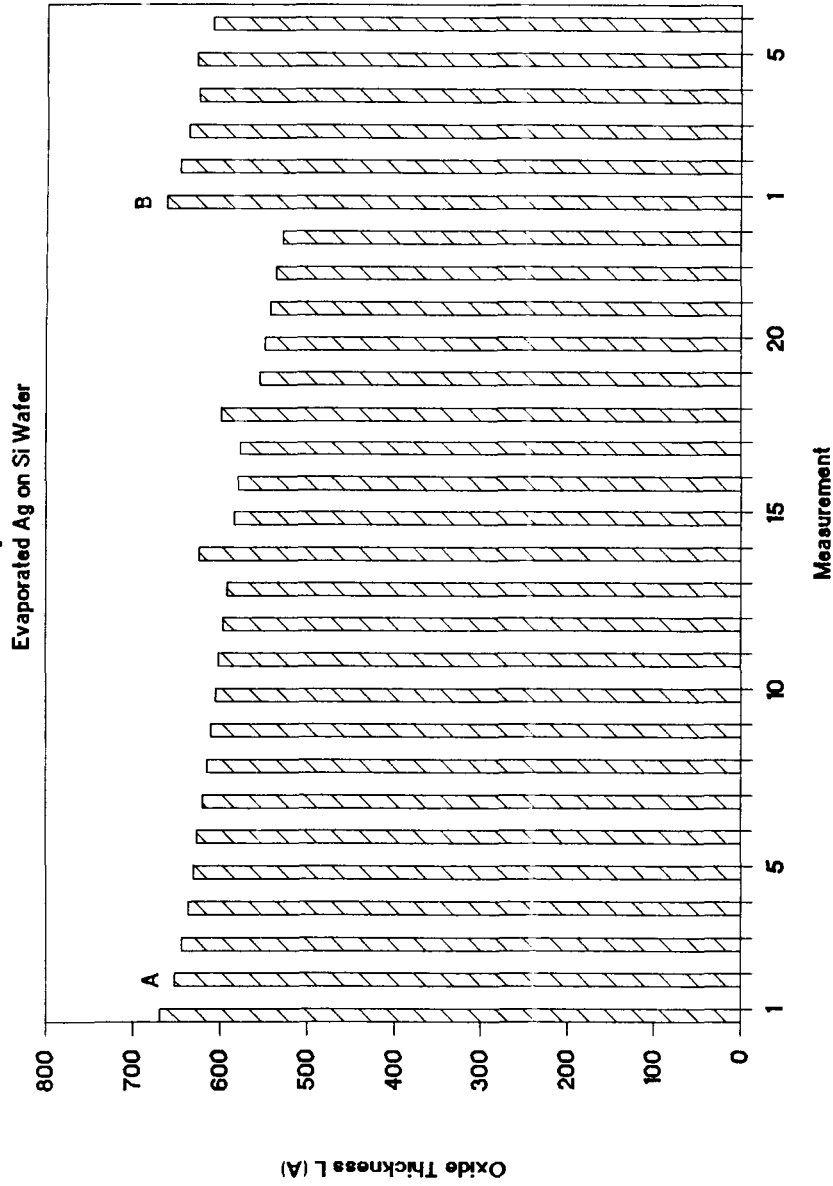


Figure 1 - Curious Heisenberg uncertainty interaction of the measurement beam of the ellipsometer with the measured oxide film.

Oxide Kinetics: Silver at T = 20 C

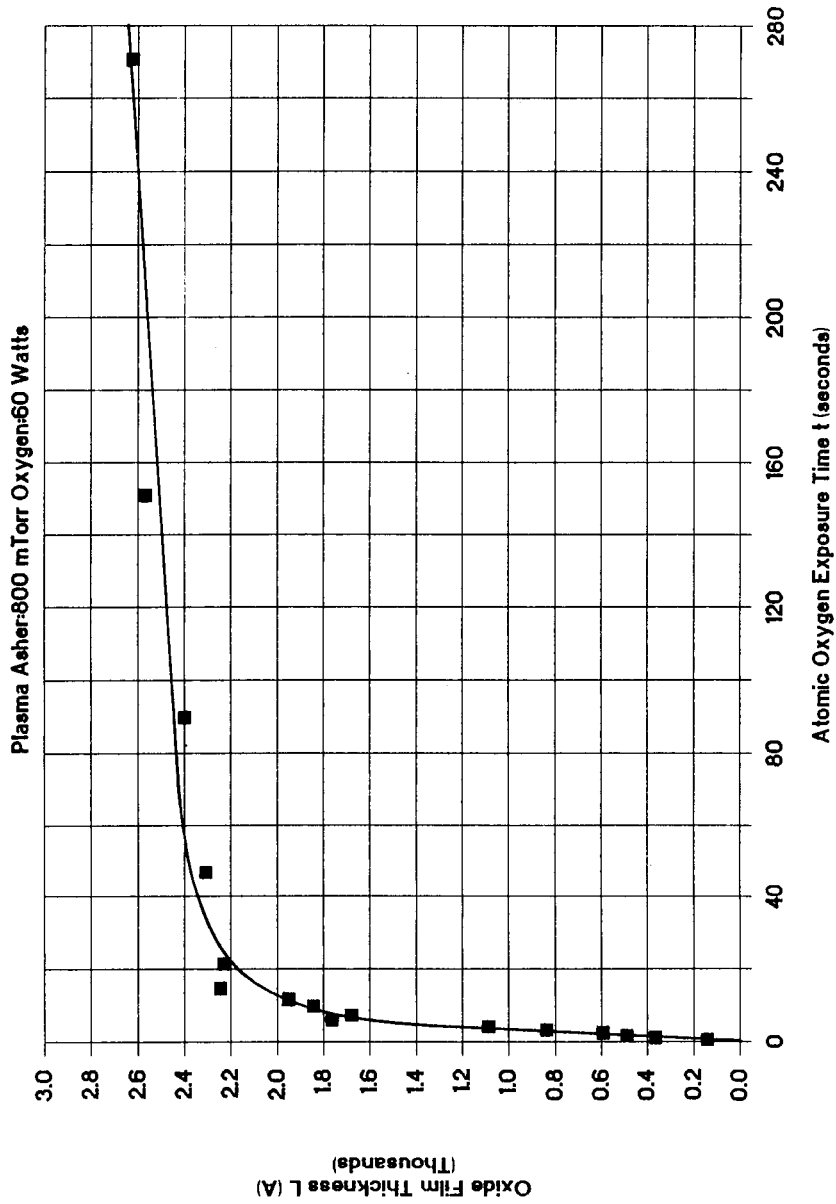


Figure 2 - Linear plot of the oxide film thickness versus exposure time for a temperature of 20 C.

Oxide Kinetics: Silver at T = 40 C

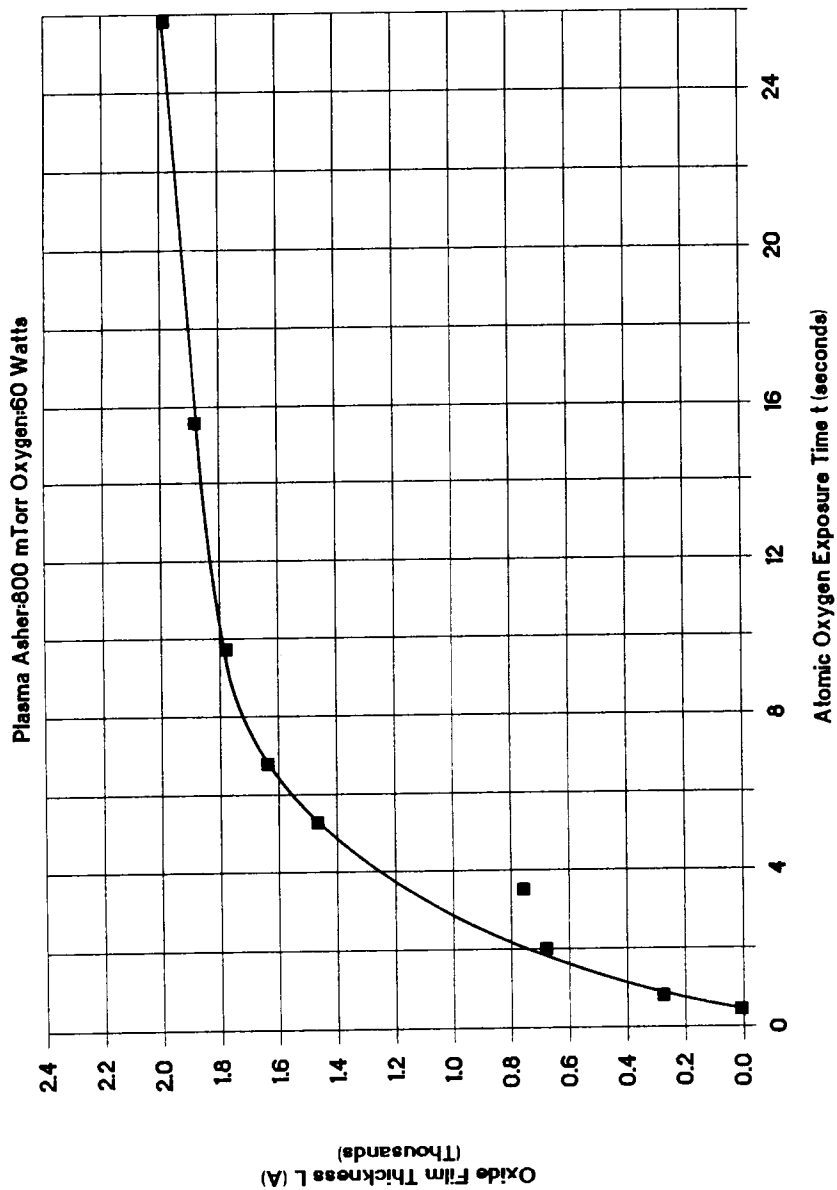


Figure 3 - Linear plot of the oxide film thickness versus exposure time for a temperature of 40 C.

Figure 4 illustrating corresponding data taken at 60°C also exhibits the same behavior typical of Figs. 2 and 3, namely a fast growth period followed by a slow growth period. For an exposure of 45 seconds, the oxide thickness achieved can be noted to be approximately 2100 Å.

Data taken at the other temperatures of this study were found to exhibit the same general trends illustrated by Figs. 2-4. These data are not specifically shown herein, but the results are used in our analysis. There was a general trend for the thickness at an arbitrary exposure time to increase with increasing temperature, indicating a thermally-activated oxide growth process.

D. Related Observations

Examination of the oxidized surface under a low power microscope showed that black islands form and increase in size in the later stages of exposure. Furthermore, a point is finally reached after extensive exposure to the atomic oxygen where the integrity of the oxide film is destroyed. At that point, the surface of the oxide when observed visually appeared to be flakey. These observations are particularly relevant to the lack of protectiveness of silver oxide for the underlying silver under attack by atomic oxygen. It may be stated that even though the growth kinetics observed for the early stages of oxidation in this study would seem to imply that the oxide layer being formed represents a highly protective barrier layer in preventing continued reaction of oxygen and silver, as reflected in the slow growth rate found in the later stage of growth, this is a bit misleading because this phase of oxidation is followed by a second phase of the reaction which is very deleterious to the coherence of the layer.

Lack of uniqueness in interpreting ellipsometer readings for nonhomogeneous oxide films makes it advisable to complement those results with data obtained by other techniques. The most direct method we used to verify that the thickness of the oxidized silver is increased over that of the pure silver was to trace the diamond stylus of a mechanical surface profiler from the bare silicon wafer onto the surface of the silver layer before reaction and after reaction. The increase in volume produced by oxidation of silver is a factor of approximately 1.5, so the oxidized silver is thicker. A third experimental probe which we used to examine some of our oxidized specimens was that of Rutherford backscattering, viz., RBS. The spectra were generally consistent with the ellipsometer deductions.

A good indication that the oxide film constitutes a separate phase on the surface of the silver was given by the interference colors observed on the silver specimens after various times of exposure to the atomic oxygen. Light gold was followed by a deeper gold, subsequently followed by violet and then blue. Oxidizing further gave a silver color, followed by second-order interference colors beginning with light gold.

IV. ANALYSIS

A. Growth Law

The most commonly-observed thermally activated oxide film growth law (4,5) is that in which the film thickness increases as the square root of the exposure time. This is known as the **parabolic growth law**. Theoretically such a law follows, for example, from the application of Fick's laws of diffusion to the steady-state transport of uncharged point defect species of oxygen or metal atoms moving through the barrier represented by the existing oxide film. Once the diffusing species reach the other side of the oxide film, the chemical reaction of new oxide formation can occur, and the oxide film is increased in thickness. The parabolic law also holds under similar conditions whenever the moving point defect species are charged. Therefore the theoretical base is strong for a parabolic growth law of some type, in which case

$$L(t) \propto t^{1/2}, \quad (1)$$

where $L(t)$ is the oxide thickness at time t . The proportionality constant appropriate for this relation would depend upon the specifics of the experimental parameters and the metal, the state of charge of the mobile point defect species, the mobilities and boundary concentrations of the species involved, and the amount of space charge in the oxide due to these mobile species (4,5). Independent of the actual value of the proportionality constant, however, plots of the oxide film thickness versus the square root of the exposure time would be straight lines if any type of parabolic growth law is applicable, irrespective of the specific value of the rate constant.

To test whether our experimental data conform to a parabolic growth law, we made plots at each temperature of the oxide thickness versus the square root of the exposure time. One example of these plots is shown in Fig. 5, where the Fig. 4 data taken at 60°C are replotted in parabolic form. It is clear from a cursory inspection of Fig. 5 that a straight-line relation is not obtained for the data by such a plot, so that a parabolic growth law is definitely not appropriate for describing our data. The parabolic plots for the data taken at other temperatures in this experimental study likewise gave no support for the applicability of the parabolic growth law.

An alternate growth law sometimes found to fit empirically some experimental oxide growth rate data is the cubic growth law, for which the oxide film thickness increases as the cube root of the time,

$$L(t) \propto t^{1/3}. \quad (2)$$

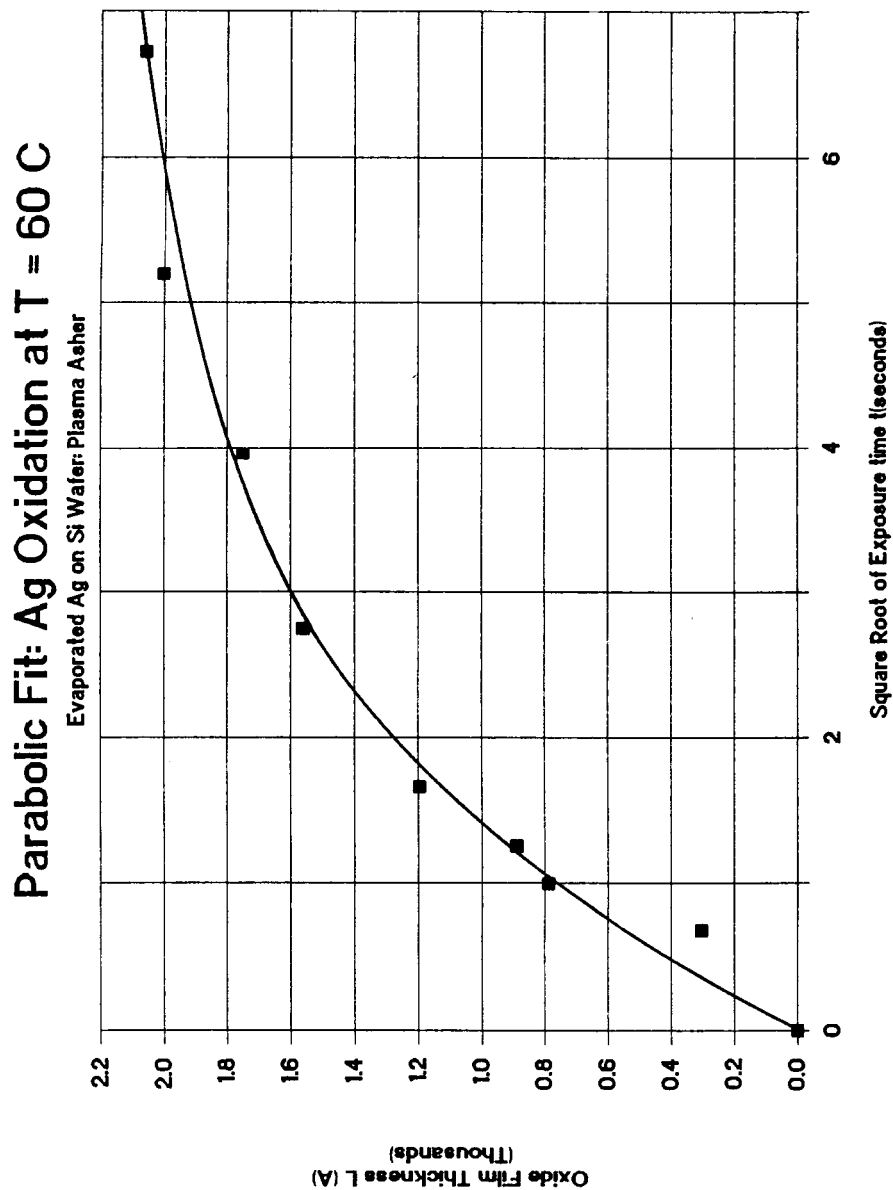


Figure 5 - Parabolic plot of the oxide film thickness versus exposure time for a temperature of 60 C.

To explore this possibility, we made plots of the thickness as a function of the cube root of the time for our data at each of the temperatures of this study. Figure 6 shows the plot for the 60°C data. It is evident that a straight-line relation does not describe well the experimental data plotted in this manner. The corresponding plots at the other temperatures of this study likewise did not obey straight-line relationships, so the conclusion was reached that our data do not conform to a cubic growth law.

The empirical logarithmic growth law

$$L(t) \propto \log t \quad (3)$$

is sometimes found to be useful for fitting experimental oxidation data. Because our data failed to conform to both the theoretically well-based parabolic growth law and the more empirical cubic growth law, the next step was to try semilogarithmic plots of the data to see if the empirical logarithmic law could provide a framework for describing the data.

Figure 7 shows the results of fitting a straight line to the oxide film thickness data at 60°C (cf. Fig. 4) after replotting as a function of the logarithm (to the base 10) of the exposure time. Although the straight line drawn through the data points in Fig. 7 is not by any means in perfect agreement with all of the data, it nevertheless provides an acceptable fit in the sense of giving the proper general trend. Such plots were also constructed for the data taken at the other temperatures of this study. For example, Fig. 8 shows the semilogarithmic plot of the 40°C data. Again the fit of the straight line to the kinetics data is acceptable. Since this conformity was also found for the data at the other temperatures, we concluded that the logarithmic growth law provides an acceptable description of our kinetic studies on the oxidation of silver by atomic oxygen over the temperature range 0 to 70°C.

B. Logarithmic Growth Law Parameters

The general equation for a straight line, namely $y(x) = mx + b$, where the dependent variable y changes linearly with the independent variable x , yields a line having slope m and y -intercept b . Our analog of this relation is

$$L(t) = L_{crit} \log_{10} t + b_0, \quad (4)$$

where $L(t)$ is the oxide film thickness after the accumulated exposure time t , with the parameters L_{crit} and b_0 representing the slope and the intercept, respectively. Equation (4) was fit to our data at the various temperatures. A pictorial representation of the deduced values for the slope L_{crit} at each temperature is shown in the bar graph of Fig. 9, while the deduced value of the intercept b_0 at each of the

Cubic Fit: Ag Oxidation at T=60 C

Evaporated Ag on Si Wafer: Plasma Asher

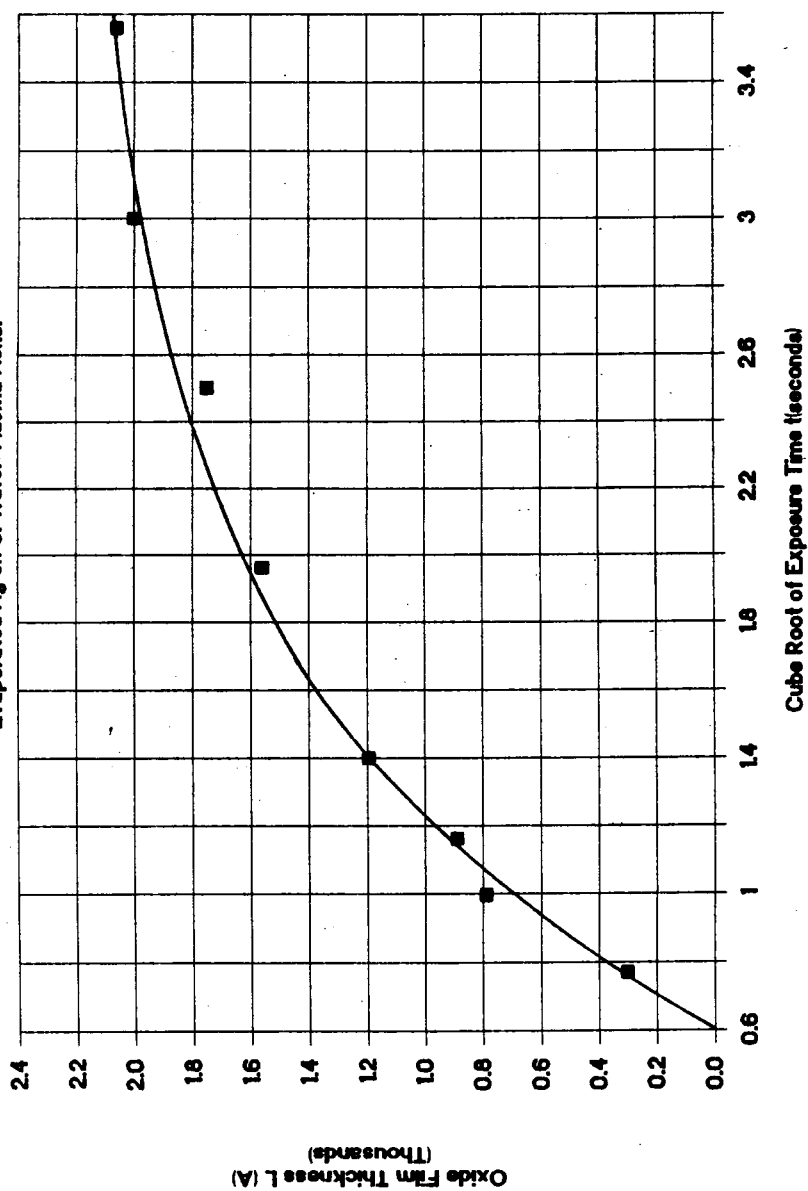


Figure 6 - Cubic plot of the oxide film thickness versus exposure time for a temperature of 60 C.

Logarithmic Growth: Ag Oxidation at 60 C

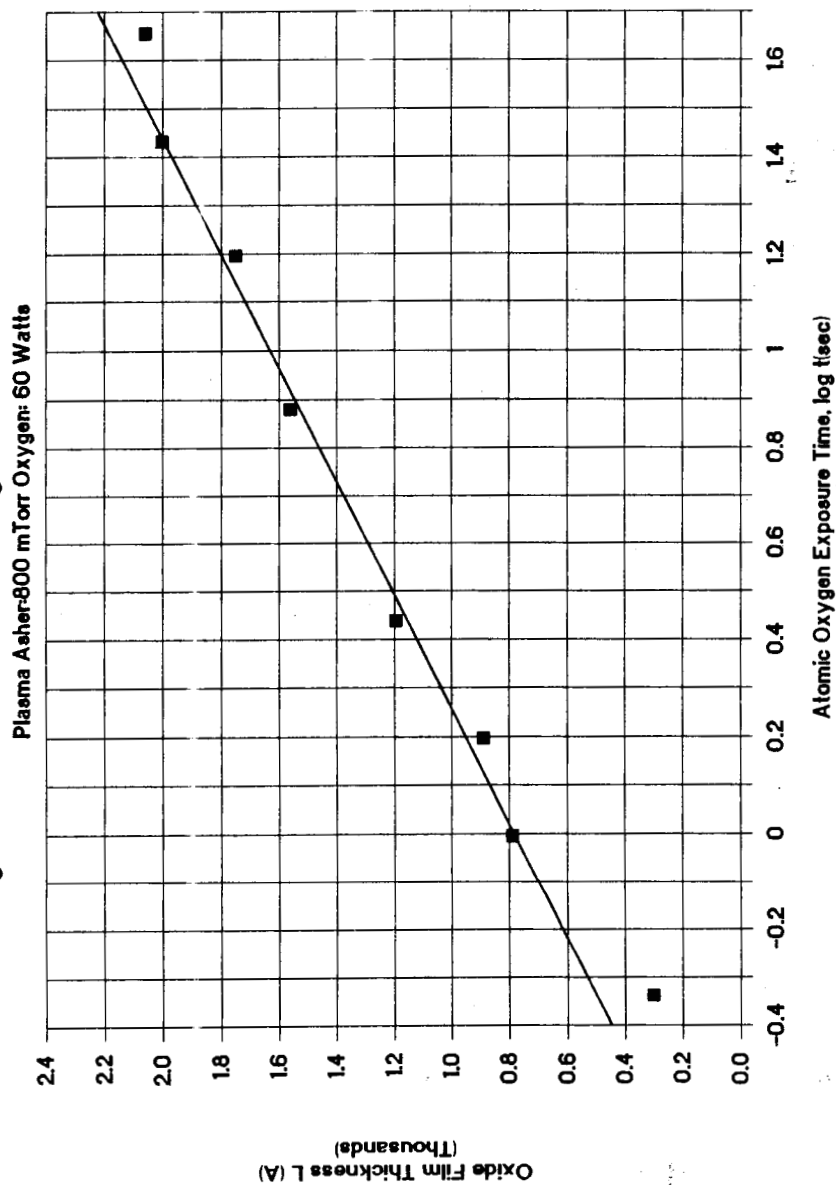


Figure 7 - Semilogarithmic plot of the oxide film thickness versus exposure time for a temperature of 60 C.

Logarithmic Growth:Ag Oxidation at 40 C

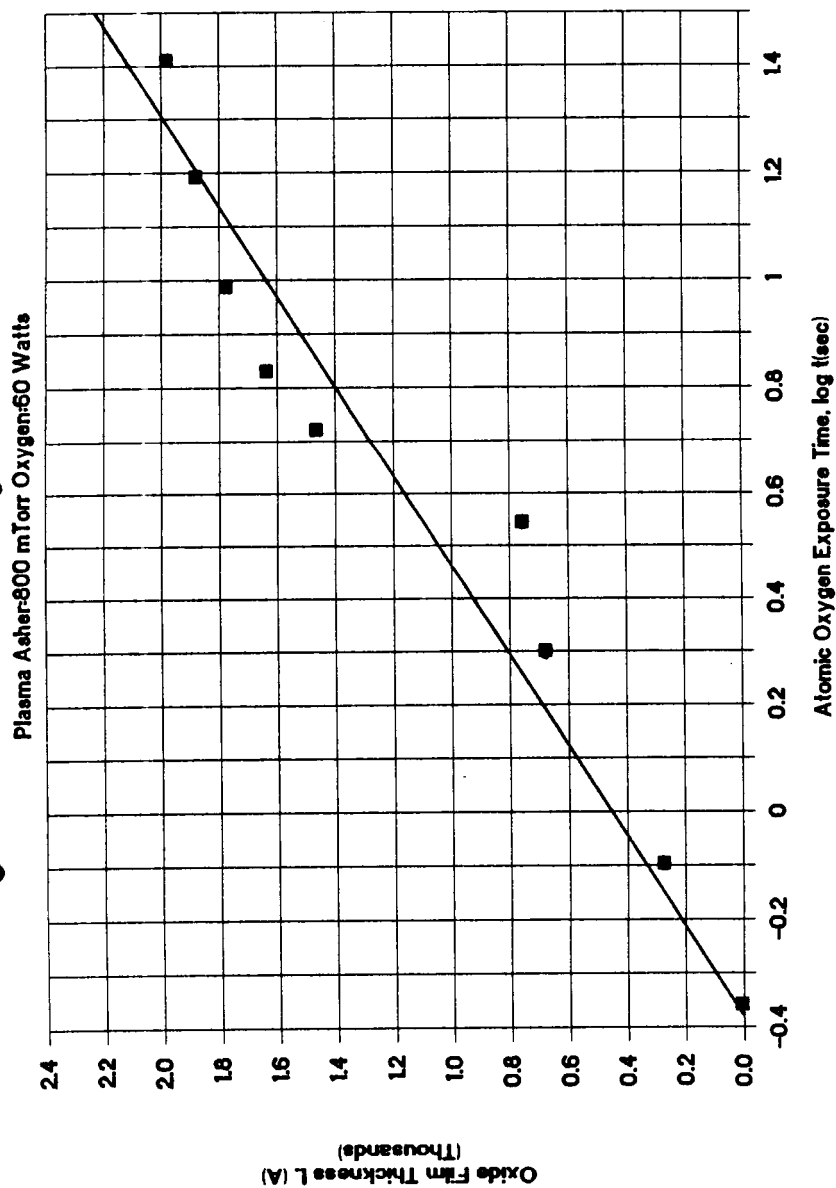


Figure 8 - Semilogarithmic plot of the oxide film thickness versus exposure time for a temperature of 40 C.

Oxidation of Ag on Si in Plasmod Asher

Logarithmic Kinetics Temperature Effect

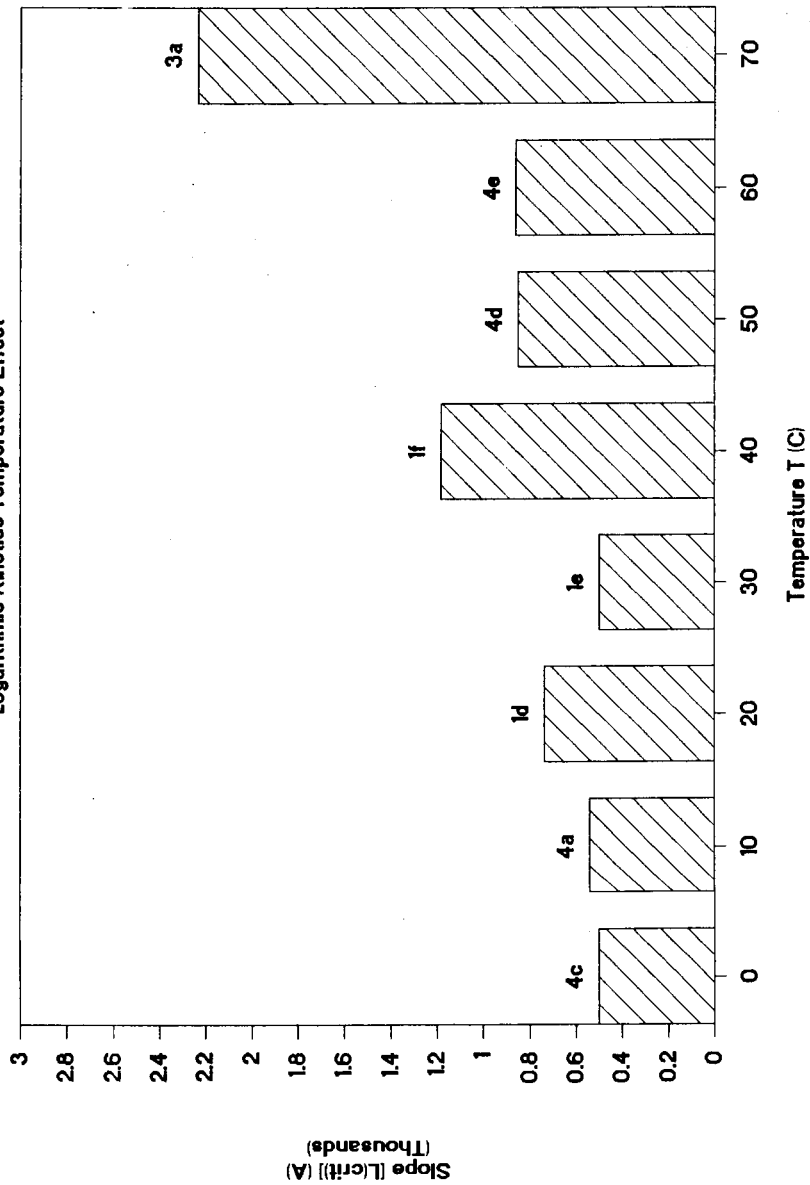


Figure 9 - Temperature variation of the slope of the logarithmic plot of the kinetics.

temperatures is shown in the bar graph of Fig. 10. Note the tendency for both L_{crit} and b_0 to increase with increasing temperature.

Without adopting a specific microscopic model yielding a theoretically well-founded logarithmic growth law, it is not possible to predict the temperature-dependence of the parameters L_{crit} and b_0 which appear in Eq. (4). However the experimental trend is for these parameters to increase with increasing temperature, as noted in the bar graphs of Figs. 9 and 10. This suggests constructing Arrhenius plots for these parameters.

An Arrhenius plot for a thermally-activated quantity $Z(T)$ yields a straight-line relation when the natural logarithm $\ln[Z(T)]$ is plotted as a function of the reciprocal of the absolute temperature T . For our parameters L_{crit} and b_0 , this would correspond to the functional dependences

$$(5) \quad L_{crit} \propto \exp(-U_C/k_B T),$$

$$(6) \quad b_0 \propto \exp(-U_B/k_B T),$$

where k_B is the Boltzmann constant and U_C and U_B are parameters having units of energy. The parameters U_C and U_B are related in some presently unspecified way to the activation energy U_{act} for the reaction. The relationship may be half the activation energy U_{act} , or some other function of U_{act} , depending upon the particular microscopic theory underlying the logarithmic law [Eq. (4)] for this particular experimental system.

Plots were constructed for $\ln(L_{crit})$ versus $1/k_B T$ and for $\ln(b_0)$ versus $1/k_B T$, using our deduced values for L_{crit} and b_0 . The results of these plots are shown respectively in Figs. 11 and 12. The straight line fits to the data obtained in this manner, although not impressive in quality because of the relatively large scatter in the deduced parameters at the various temperatures, nonetheless provide reasonable descriptions of the trends indicated by the data. The slopes of these lines yield the following energy values:

$$U_C = 0.110 \text{ eV},$$

$$U_B = 0.137 \text{ eV}.$$

The precision of these values is certainly no better than 50%. If one wishes to be more conservative, taking into account the rather limited temperature range studied, the uncertainty in the cumulative exposure times due to the extremely short incremental exposure periods, and the relatively large scatter in the experimental data, then one should probably not claim the precision to be better than a factor of 2.

Oxidation of Ag on Si in Plasmod Asher

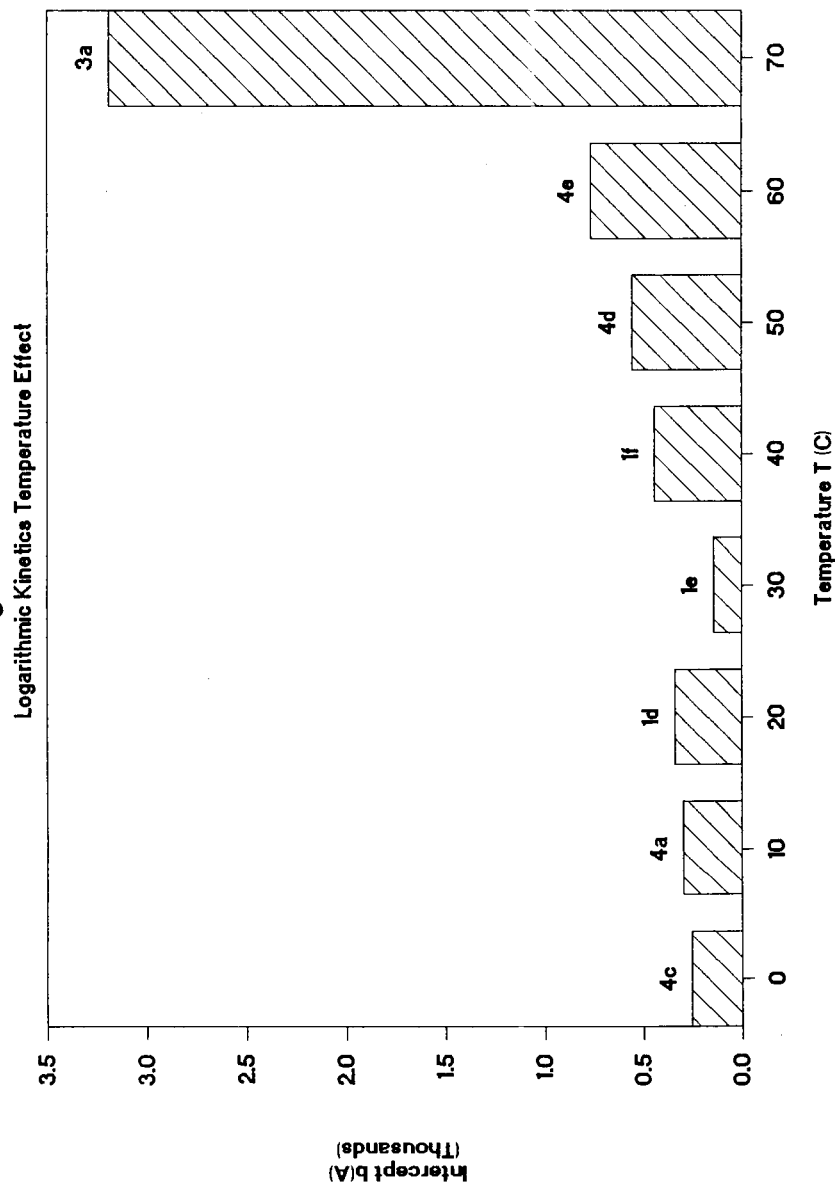


Figure 10 - Temperature variation of the intercept of the logarithmic plot of the kinetics.

Logarithmic Kinetics Slope

Oxidation of Silver in Atomic Oxygen

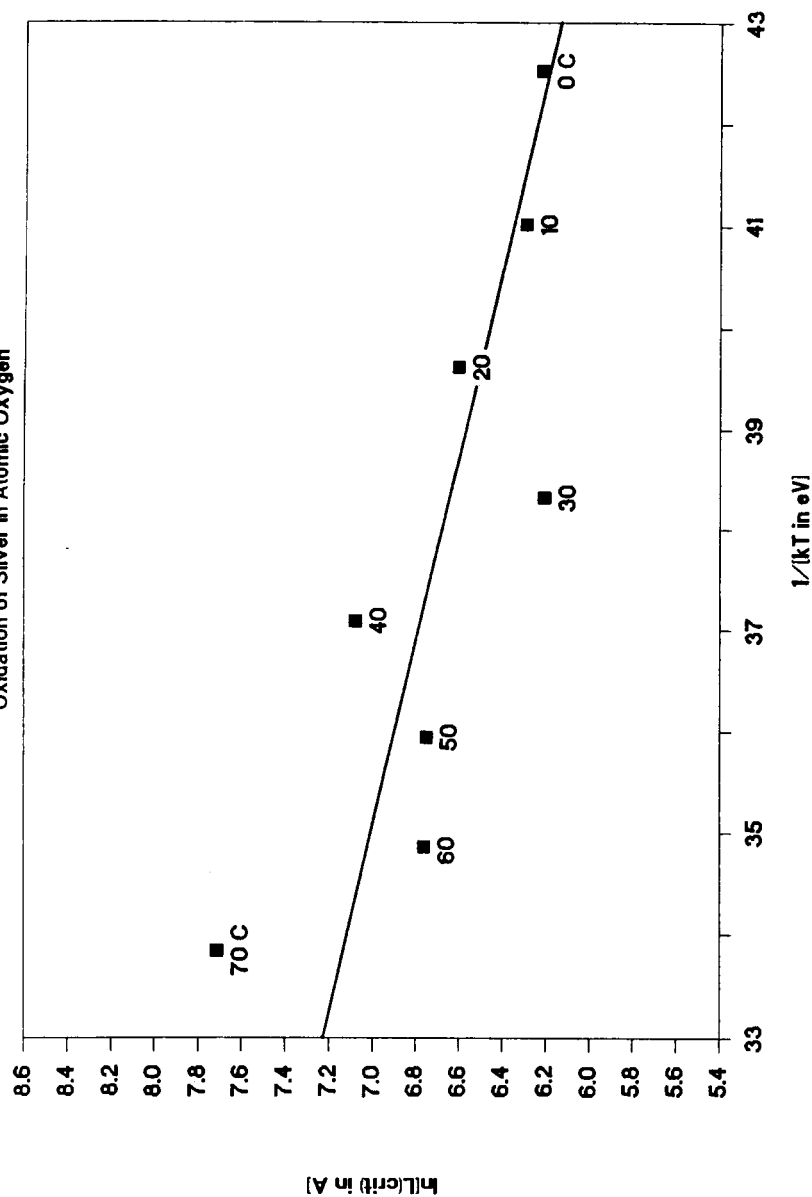


Figure 11 - Arrhenius plot of the slope for the logarithmic kinetics.

Logarithmic Kinetics Intercept

Oxidation of Silver in Atomic Oxygen

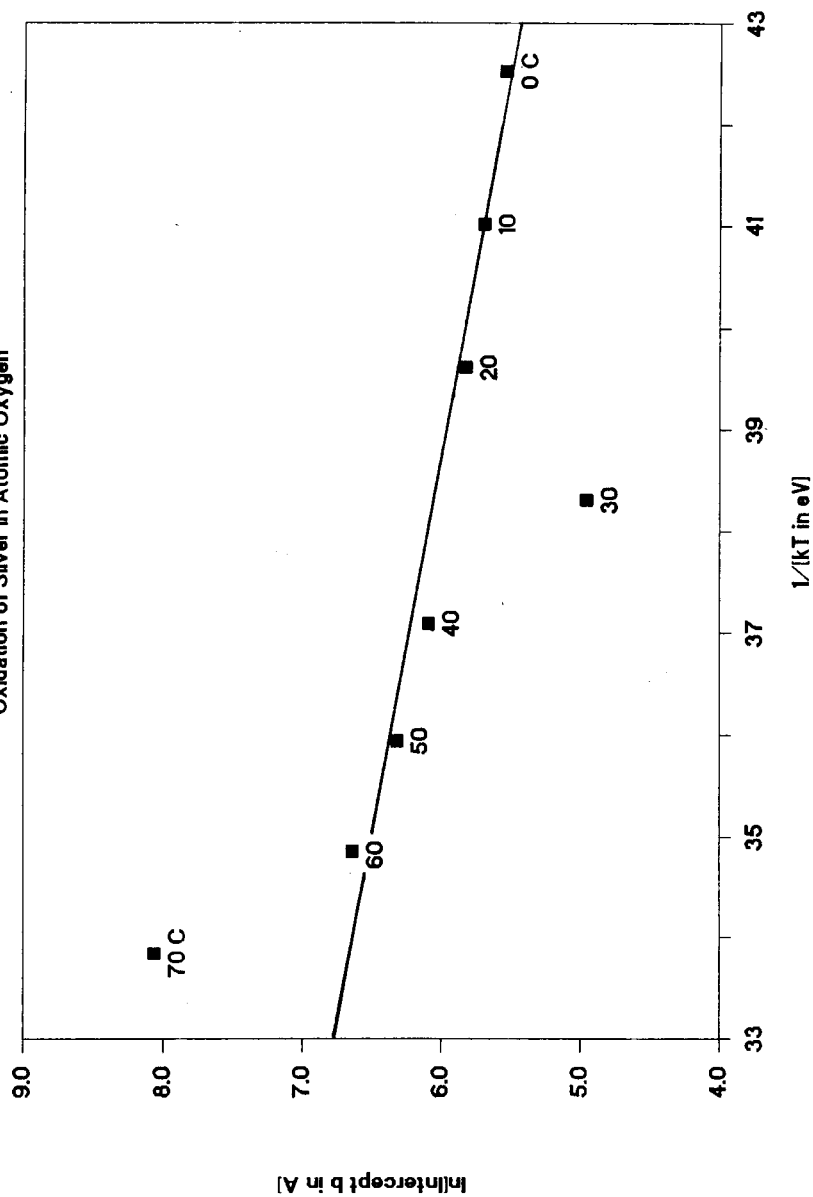


Figure 12 - Arrhenius plot of the intercept for the logarithmic kinetics.

V. CONCLUSIONS

1. Early stage oxidation of silver in atomic oxygen occurs in accordance with the logarithmic growth law.
2. The oxidation is thermally activated, the rate increasing somewhat with increasing temperature.
3. There is an interesting interaction of the laser beam with the oxide.

VI. RECOMMENDATIONS

My first recommendation is to design and construct a reaction chamber which would fit directly in the ellipsometer so that measurements could be carried out continuously *in situ*. This would eliminate a measure of uncertainty introduced by truncating the exposure periodically to carry out measurements of the oxide film thickness with the ellipsometer. Next, discharge conditions (power and oxygen pressure) should be sought which lead to a much slower reaction rate than that encountered in the present experimental study. The reason for this is that continuous exposure under the present experimental conditions would lead to growth rates which could not be followed by the automated ellipsometer even using the instrument in its most optimum fast mode. Then software should be developed to compute the oxide film thicknesses *off-line* after the measurements are complete because the process of varying optical constants until good convergence is attained *on-line* would be impossible if the film were itself changing rapidly during measurement.

A different type of recommendation is to extend the present study to bulk silver specimens to see how the results are affected by the additional roughness and different texture of mechanically-polished surfaces, relative to smooth evaporated surfaces. The refractive index of evaporated silver is about 0.1, while that of mechanically-polished bulk silver surfaces is frequently found to be in the range 0.3 to 0.6.

The effects of grain boundaries on the silver reaction with atomic oxygen are also of great interest. A comparison of data taken on single crystals, which have no grain boundaries, with corresponding data taken on bulk samples, would thus be valuable. At present it is not known to what extent the atomic oxygen diffuses into silver by bulk diffusion and to what extent the diffusion is enhanced by the presence of grain boundaries.

More attention should also be given to theoretical aspects. Assuming that the early-stage logarithmic growth law deduced in this study is confirmed by more precise and extensive experiments carried out over a wider temperature range, then it would be of great interest to seek the basic mechanism which explains such growth. In addition, the effects of mechanical stress should be included in the theoretical formulation.

Finally, these studies should be extended to other metals. It would be relatively easy, for example, to duplicate this series of experiments for evaporated films of aluminum, copper, nickel, and gold. In addition, measurements should be made on elemental metals and alloys which are potentially important for the forthcoming space station.

REFERENCES

1. A. F. Whitaker, S. A. Little, R. J. Harwell, D. B. Griner, R. F. DeHaye, and A. T. Fromhold, Jr., "Orbital Atomic Oxygen Effects on Thermal Control and Optical Materials -- STS-8 Results", AIAA-85-0416, AIAA 23rd Aerospace Sciences Meeting, January 14-17, 1985, Reno, Nevada.
2. Albert T. Fromhold, Jr., Ann F. Whitaker, Kasra Daneshvar, and Sally A. Little, "Reaction of Metals in Lower Earth Orbit during Space Shuttle Flight 41-G (STS-17)", AIAA-86 Aerospace Sciences Meeting held in Houston, Texas, November, 1986.
3. A. F. Whitaker, private communication, July, 1988.
4. A. T. Fromhold, Jr., *Theory of Metal Oxidation. Vol. I -- Fundamentals* (North Holland Publishing Company, Amsterdam, 1976), pp. 104, 135, 146.
5. A. T. Fromhold, Jr., *Theory of Metal Oxidation. Vol. II -- Space Charge* (North Holland Publishing Company, Amsterdam, 1980), pp. 178-179, 188.

N89-21739

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

CONSTRUCTION OF AN ALTERNATING GRADIENT
MAGNETOMETER

Prepared by:	Michael M. Garland
Academic Rank:	Professor
University and Department:	Memphis State University Department of Physics
NASA/MSFC:	
Laboratory:	Space Science
Division:	Astrophysics
Branch:	Cryogenic Physics
MSFC Colleague:	Eugene W. Urban
Date:	August 19, 1988
Contract No.:	NGT01-002-099 University of Alabama

CONSTRUCTION OF AN ALTERNATING GRADIENT
MAGNETOMETER

by

Michael M. Garland
Professor of Physics
Memphis State University
Memphis, Tennessee

ABSTRACT

A magnetometer is described which was constructed to facilitate the study and characterization of the magnetic properties of high transition temperature superconductors. This instrument has been used to measure the D.C. magnetic susceptibility of several superconducting compounds as a function of temperature.

The construction of the magnetometer and the operating parameters are discussed in detail.

ACKNOWLEDGEMENTS

I would like to thank Dr. Eugene Urban, Dr. Palmer Peters, Mr. Charles Sisk and Mr. James Jolley for the assistance and encouragement which they have provided during the course of this project.

I also wish to thank the ASEE and NASA for providing the opportunity for me to work at MSFC.

LIST OF FIGURES

1. Magnetometer System	XIII-6
2. Detail of Vane and Coils	XIII-7
3a. Detail of Vane	XIII-9
3b. Detail of Sweep Coils	XIII-9
4. Field Map of Sweep Coils	XIII-10
5. Field Map of Field Coils	XIII-11
6. Data for SM123 + AgO 3:1	XIII-14
7. X_m vs T for SM123 + AgO 3:1	XIII-15
8. X_m vs T for $T > T_c$	XIII-16
9. $1/X_m$ vs T for $T > T_c$	XIII-17
10. $1/X_m$ vs T	XIII-18

INTRODUCTION

The magnetic properties of materials form the basis of many practical applications. For this reason the study of these properties is an important part of modern solid state physics. The response of a solid to an external magnetic field is described by the equation

$$B = H + 4\pi M \quad (\text{cgs units}) \quad (1)$$

H is the strength of the applied field, in Oersteds. B is the magnetic induction within the material, in Gauss, and M is the magnetization. M is a property of the material and in a homogeneous solid it is equal to the average magnetic moment per unit volume. In paramagnetic and diamagnetic homogeneous solids M is directly proportional to H so

$$M = X_m H \quad (2)$$

where X_m is the magnetic susceptibility. In a paramagnetic substance X_m is positive and in a diamagnetic substance it is negative. For simple homogeneous materials a graph of M vs H yields a straight line of slope X_m .

The alternating gradient magnetometer operates on the principal that a material having magnetization M will experience a force when placed in a spatially varying magnetic field.

$$F = M (dH/dx) \quad (3)$$

M is the component of magnetization in the x-direction and dH/dx is the x-component of the magnetic field gradient. The construction of a gradient magnetometer using an alternating field

$$H = H_0 \sin \omega t \quad (4)$$

has been described by P. J. Flanders¹ The instrument described here is based on that paper.

OBJECTIVES

The objectives of this project were to:

1. Construct an alternating gradient magnetometer capable of performing magnetic susceptibility measurements on small solid samples.
2. Measure the D.C. magnetic susceptibility of high transition temperature superconductors as a function of temperature from room to liquid nitrogen.

THEORY

I. The Magnetometer

An alternating magnetic field of the form $H = H_0 \sin \omega t$, where H_0 has a constant gradient dH_0/dx , is produced by a pair of sweep coils. The sample, whose magnetization is to be measured, is placed on the end of a vane which forms a cantilever. The alternating field produces a sinusoidal force on the sample which causes it, and the cantilever, to vibrate at the frequency ω . The upper part of the vane is a piezoelectric bimorph which produces a voltage proportional to the deflection of the sample. Since the deflection of the sample is proportional to the applied force and the force is proportional to the magnetization, the piezoelectric voltage is proportional to the magnetization of the sample. If a large direct current field H_{dc} is applied, and if H_{dc} is much larger than H_0 , then the D.C. magnetization can be measured as a function of magnetic field.

The sensitivity of the magnetometer is optimized by vibrating the vane at its normal mode frequency, or one of its overtones. Since the vane is vibrating at normal resonance it is very sensitive to external disturbances and must be isolated, as nearly as possible, from outside oscillations.

II. Superconductivity

If an ideal superconductor is placed in an external magnetic field in the x-direction the field will penetrate a small distance into the surface. The internal magnetic induction is given by

$$B = B_0 \exp(-x/L) \quad (5)$$

L , the penetration depth is of the order of 1000 \AA in pure elemental superconductors. For bulk samples the internal magnetic induction $B = 0$, or approximately so. The superconductor in this state

is a perfect diamagnet with susceptibility

$$X_m = -1/4\pi \quad (6)$$

So the magnetization $M = -H/4\pi$ in an ideal bulk superconductor. A plot of M vs H will yield a straight line having slope X_m . In granular superconductors such as the ones under study at MSFC the magnetic field will penetrate the sample and the flux lines can be "pinned" by impurities and grain boundaries. This leads to hysteresis in the M vs H curves. This effect has been observed by D. Wong et al in the case of small fields².

The recent discovery of superconductivity at temperatures above the normal boiling point of liquid nitrogen³ (77K) has resulted in a frenzy of research activity into the properties of the new materials. Researchers at the Space Science Laboratory have made significant contributions in the field^{3 4 5}, in collaboration with scientists at The University of Alabama in Huntsville and at Lockheed Missiles & Space Co. Current research has been centered around materials having a nominal composition $R_1Ba_2Cu_3O_x + AgO$ where R is a rare earth ion and x is approximately 6.8 . These materials have shown several unusual magnetic properties which warrent further investigation.

APPARATUS

1. The System

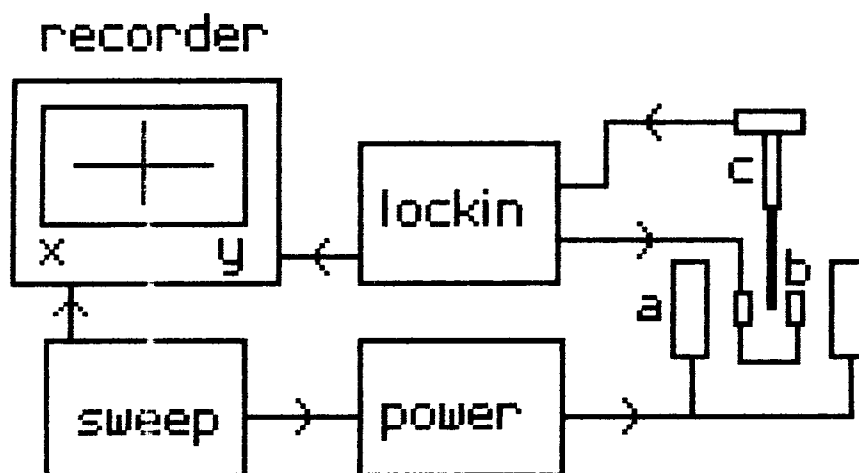
The complete system is composed of (a) Sweep Coils, (b) Field Coils to produce the D.C. field, (c) The Vane and Piezoelectric Transducer, and (d) Electronics. The system is shown in figure 1.

The output of a lockin amplifier is used to drive the sweep coils. The sweep coil current was maintained at 150 mA by a 5X power amplifier. The voltage output from the transducer is measured with the lockin amplifier which drives the y-axis of an x-y recorder. A bipolar operational amplifier is used to drive the field coils. The coil current is monitored and plotted on the x-axis of the x-y recorder. The resulting output is a graph of magnetization vs applied field. From this graph the D.C susceptibility can be determined. A more detailed view of the transducer and coils is shown in figure 2.

2. The Vane and Transducer

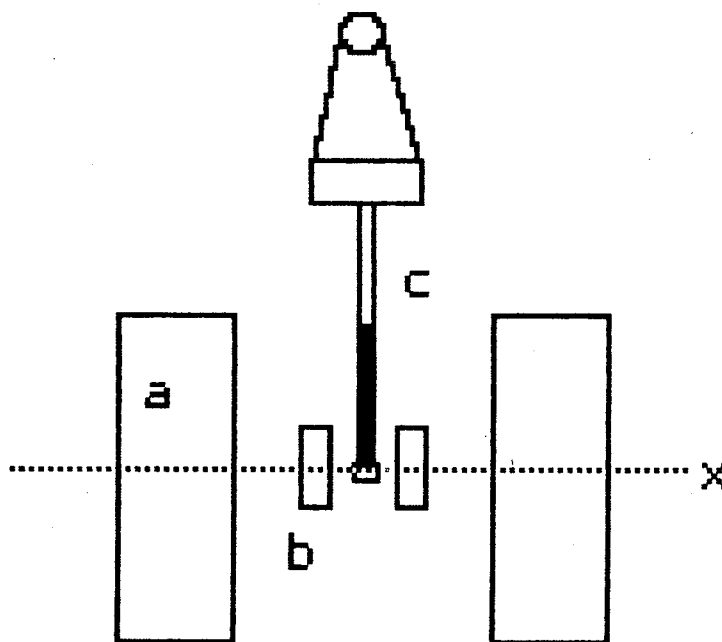
The transducer was constructed using Kynar Piezofilm⁶, a polarized polyvinylidene flouride film which was metallized on both sides with aluminum. The complete vane with transducer is shown in figure 3a. Strips of Kynar were glued to a phosphor bronze piece with silver varnish to form a piezoelectric bimorph. An extension made from 29 mil thick graphite composite was glued to the end of the bimorph with GE 7031 varnish. The vane extension allowed the sample to be cooled while the Kynar remained at or near room temperature.

The bimorph was clamped between two contacts made from copper foil and the entire assembly was suspended by two rubber bands to provide isolation from external vibrations. A #40 Copper-Constantan (Type K) thermocouple was cemented to the end of the extension with 7031 varnish so that the sample



(a) Field coils. (b) Sweep coils.
(c) vane and Transducer.

Figure 1. Magnetometer System.



(a) Field coils. (b) Sweep coils.
(c) Vane and support.

Figure 2. Detail of Vane and Coils.

temperature could be measured. The sample was attached to the end of the vane extension, in contact with the thermocouple, using silicon grease. Several vanes were constructed having resonant frequencies of 40 to 50 Hz depending on sample mass and, somewhat, on the temperature.

3. Sweep Coils

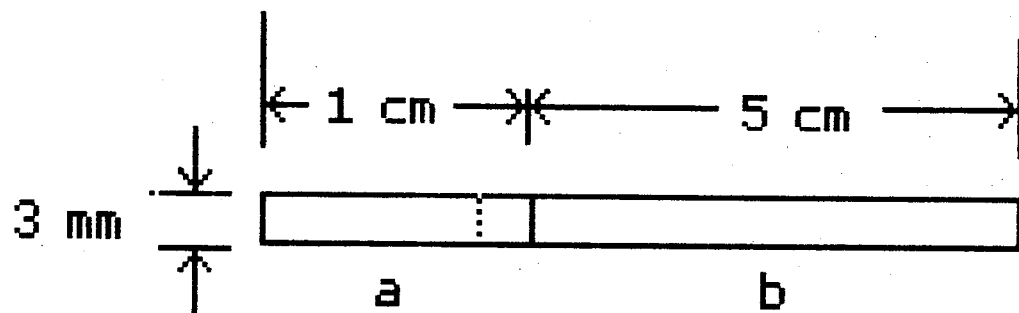
The sweep coils are shown in figure 3b. These were rigidly mounted with a separation distance of 6 mm. Figure 4 shows the magnetic field as a function of distance between the two sweep coils. The current was 100 mA D.C. and the field gradient was 22.1 Oe/cm and can be seen to be quite constant for several mm about the center position. The actual vibrational amplitude of the sample was presumed to be much less than 1 mm since the motion was not visible at any time. The sweep coils were driven at 150 mA during most of the experiments.

4. Field Coils

The field coils were air core solenoids having an outer diameter of 4.75 in, an inner diameter of 2.25 in, and a thickness of 1.50 in. Each coil had a D.C. resistance of 15 Ohms. When placed 6.5 in apart they provided a field of 26 Oe/A. Figure 5 shows the field profile for these coils. The bipolar operational amplifier used had a maximum voltage output of ± 15 V giving a maximum field of ± 52 Oe with the coils in parallel.

5. Electronics

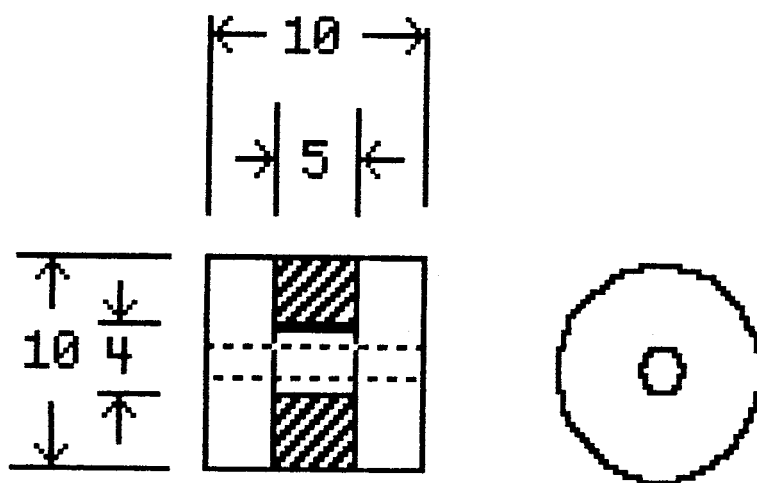
A lockin amplifier was used for synchronous detection of the transducer signal. The voltages from the two sides of the bimorph were subtracted by the amplifier. This doubled the output voltage while subtracting broadband noise caused by mechanical vibrations and electrical pickup. Much of the pickup noise could be eliminated by driving the vane slightly off resonance and measuring the quadrature (90°) signal. The output from the



(a) Piezoelectric Bimorph

(b) Vane Extension

Figure 3a. Detail of Vane.



Dimensions in mm. Wound with
250 Turns of #36 Formvar
insulated copper wire.

Figure 3b. Detail of Sweep Coils.

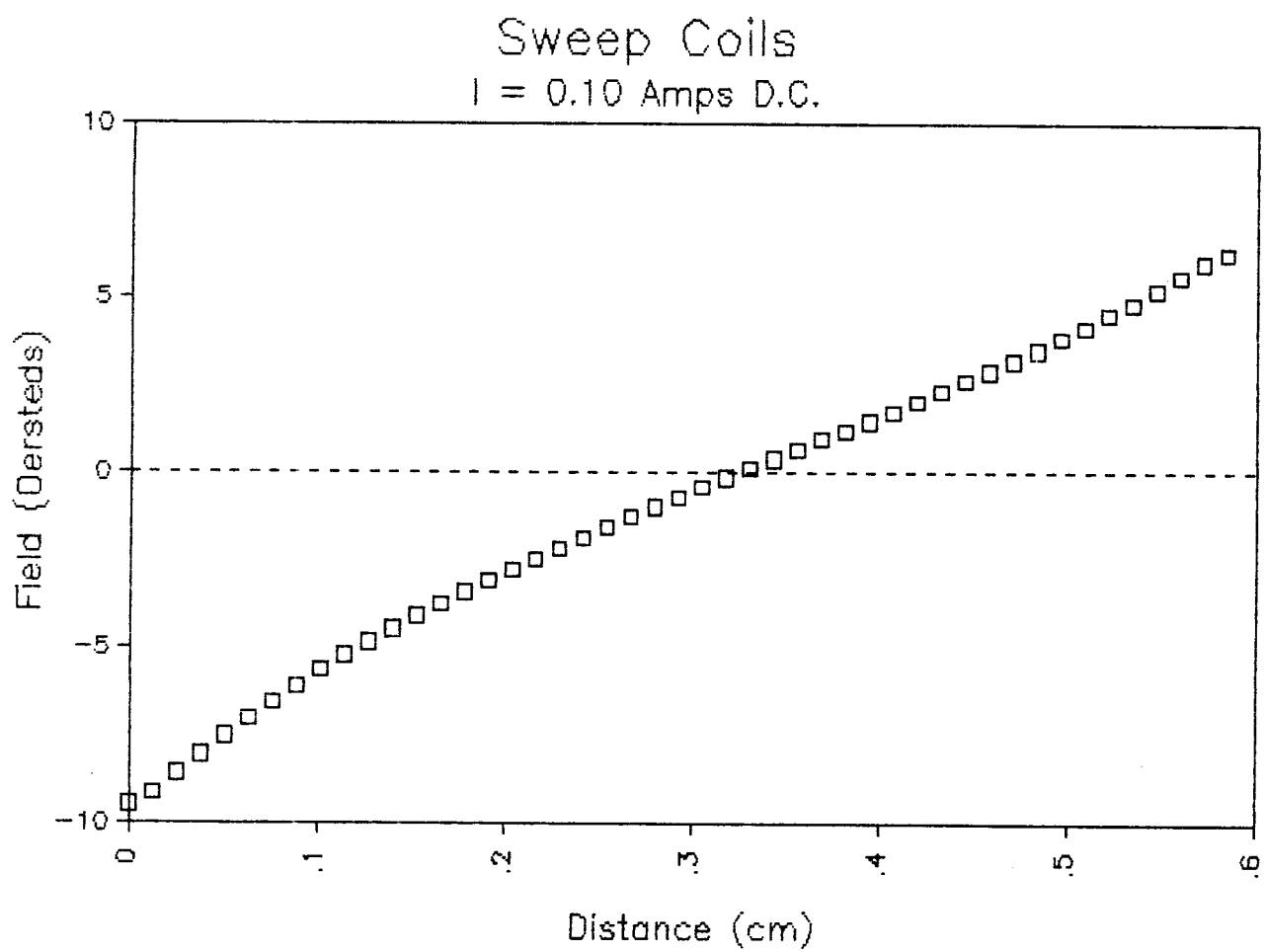


Figure 4. Field Map of Sweep Coils.

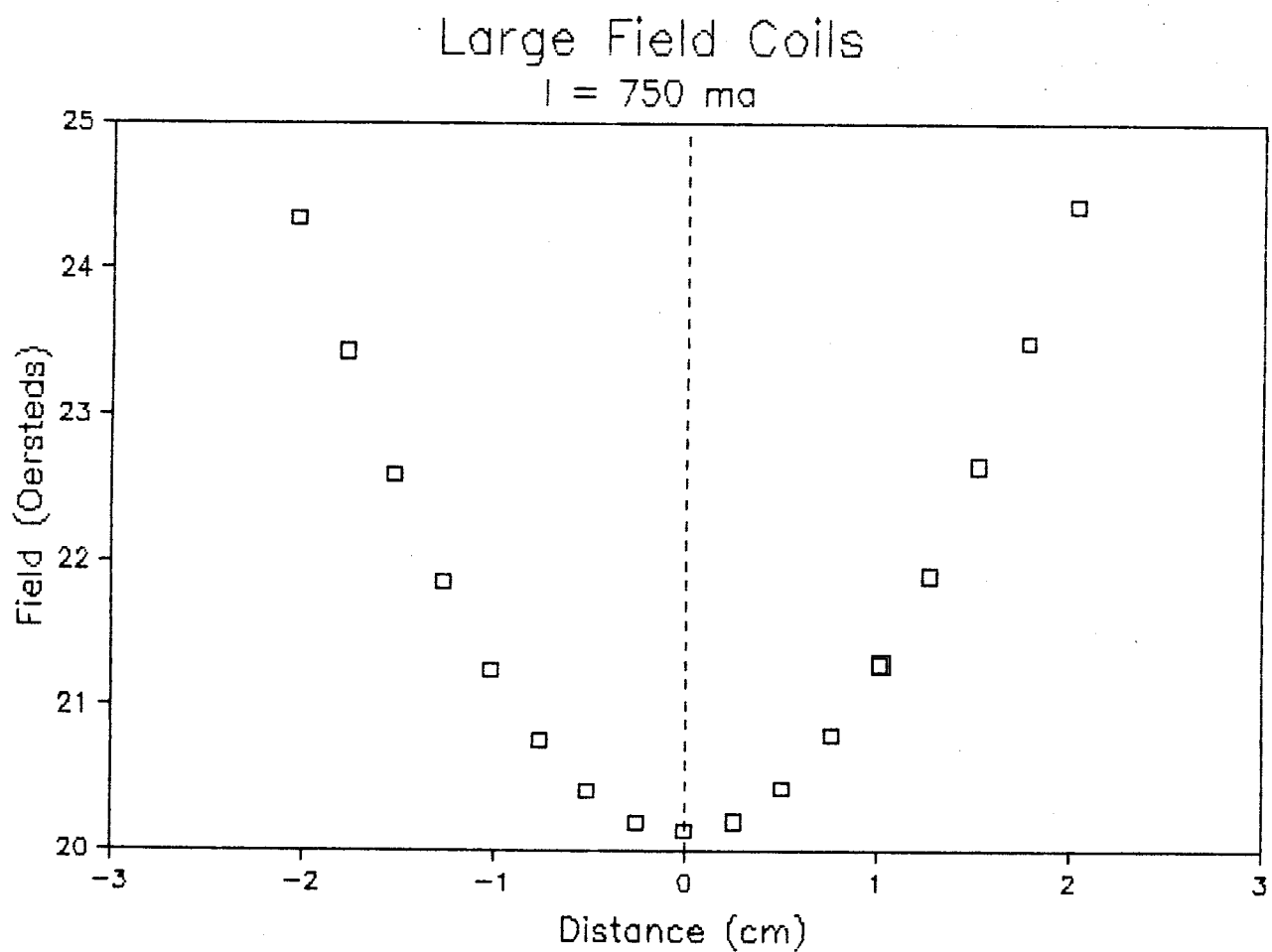


Figure 5. Field Map of Field Coils.

lockin signal channel was proportional to the sample magnetization. This was recorded on the y-axis of an x-y recorder. The field coil current was passed through a shunt resistor and the voltage from this was recorded on the x-axis of the recorder. The field coil current was swept at a rate of about 100 Oe/min.

In order to measure the magnetization as a function of temperature the sweep coil assembly was submerged in a plastic foam dewar of liquid nitrogen. The sample, on the end of the vane extension, was just above the surface of the liquid. As the liquid evaporated the sample would slowly warm up.

Operation and Experimental Results

1. Magnetometer Calibration

The magnetometer was calibrated using two samples of known susceptibility⁷. (a) A piece of Ni ribbon having a specific susceptibility of 52.38 emu/gm, and (b) A pellet of Nd_2O_3 having a specific susceptibility of 0.0102 emu/gm. Although it is ferromagnetic (X_m is a function of H) the Ni sample gave a linear M vs H curve for the low fields used. From the slope of the M vs H curves the output of the x-y recorder could be calibrated. It was found that each time a particular vane was used it needed to be recalibrated. Presumably this was due to slight changes in vane position associated with handling, which was required to change samples.

2. Thermocouple Calibration

The copper-constantan thermocouple which was used to monitor the sample temperature was calibrated at liquid nitrogen temperature and at the ice point. It was found to be off by -0.004 mV at the liquid nitrogen boiling point and +0.006 mV at the ice point. These errors correspond to 0.24 degrees C and 0.16 degrees C respectively. A thermocouple correction table was generated by taking a linear fit each 10 degrees. This should insure that the temperature measurements were accurate to within 0.1 degrees. The thermocouple correction table is given in the appendix.

3. Experimental Results

Data for a sample of $\text{Sm}_1\text{Ba}_2\text{Cu}_3\text{O}_x$ doped with AgO in a ratio of 3:1 is presented in figures 6 through 10. Since the material is paramagnetic when in the normal state and diamagnetic in the superconducting state, the superconducting transition is

ORIGINAL PAGE IS
OF POOR QUALITY

Figure 6. Data for SM123+AgO 3:1

Temp (K)	Slope (uV/A)	X (emu/gm)	1/X (gm/emu)
89.5	-5400.00	-6.1020	-.1639
90.5	-5400.00	-6.1020	-.1639
92.8	-5100.00	-5.7630	-.1735
93.1	-4800.00	-5.4240	-.1844
93.7	-4650.00	-5.2545	-.1903
94.5	-4300.00	-4.8590	-.2058
95.2	-4150.00	-4.6895	-.2132
96.7	-3800.00	-4.2940	-.2329
97.75	-3650.00	-4.1245	-.2425
98.5	-3100.00	-3.5030	-.2855
99.5	-2800.00	-3.1640	-.3161
100.1	-2250.00	-2.5425	-.3933
101.2	-1750.00	-1.9775	-.5057
101.7	-1450.00	-1.6385	-.6103
102.5	-1000.00	-1.1300	-.8850
103.6	-275.00	-.3108	-3.2180
105.4	3.25	.0037	272.2941
108.1	3.00	.0034	294.9853
110.6	2.25	.0025	393.3137
111.8	2.00	.0023	442.4779
112.7	2.00	.0023	442.4779
113.9	1.88	.0021	470.7211
116.2	1.83	.0021	483.5824

ORIGINAL PAGE IS
OF POOR QUALITY

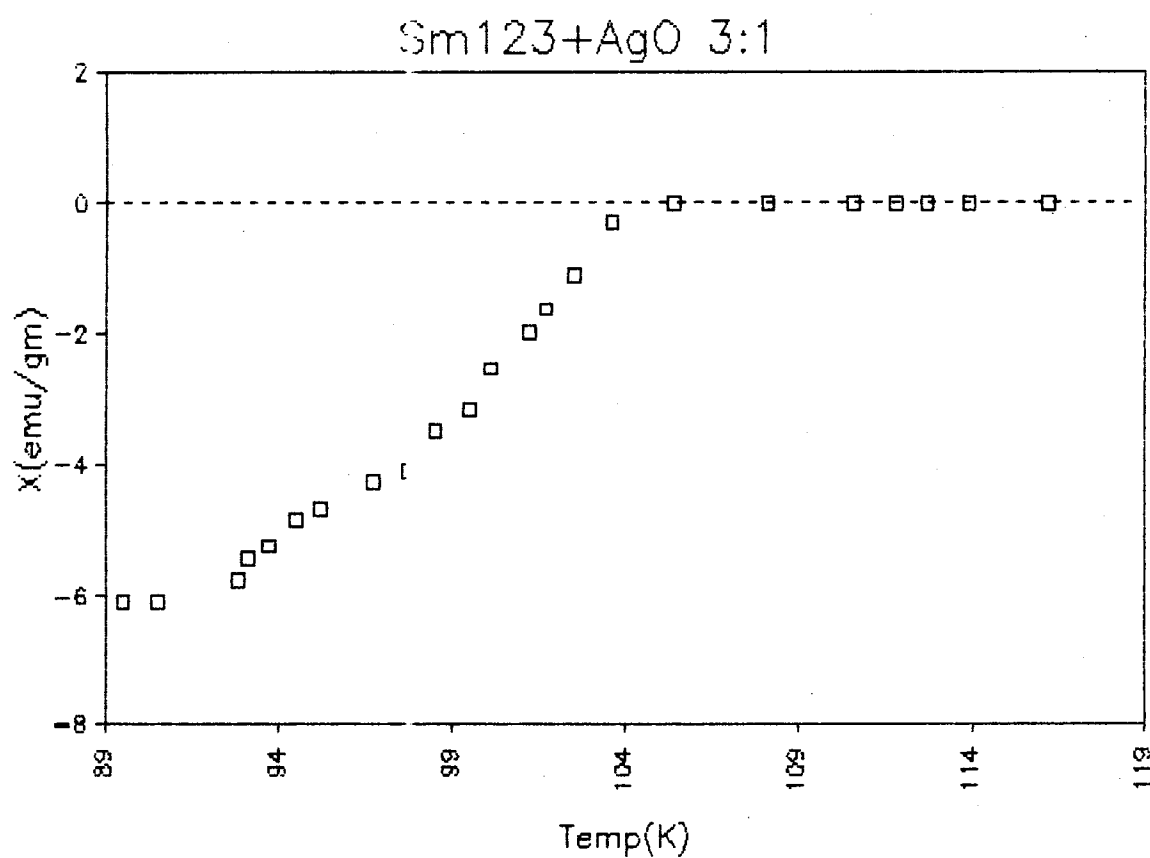


Figure 7. X_m vs T for SM123 + AgO 3:1.

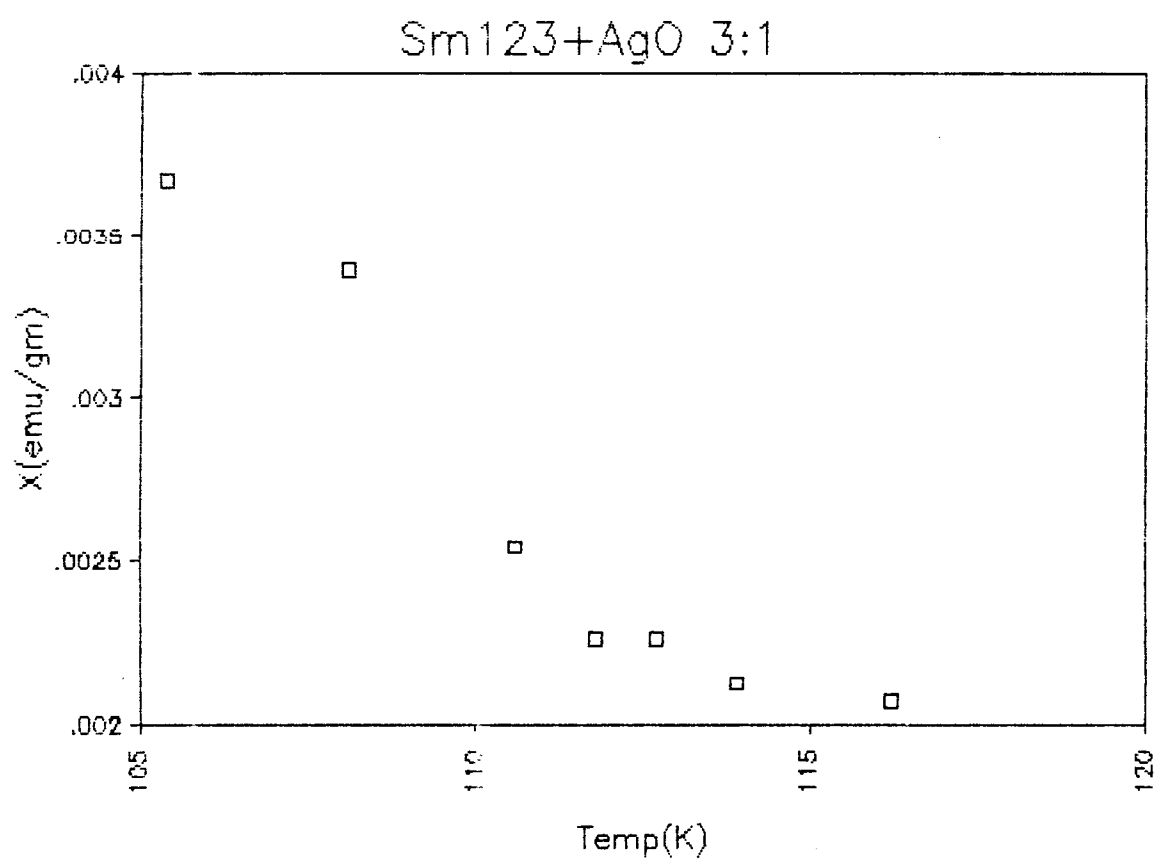


Figure 8. X_m vs T for $T > T_C$.

ORIGINAL PAGE IS
OF POOR QUALITY

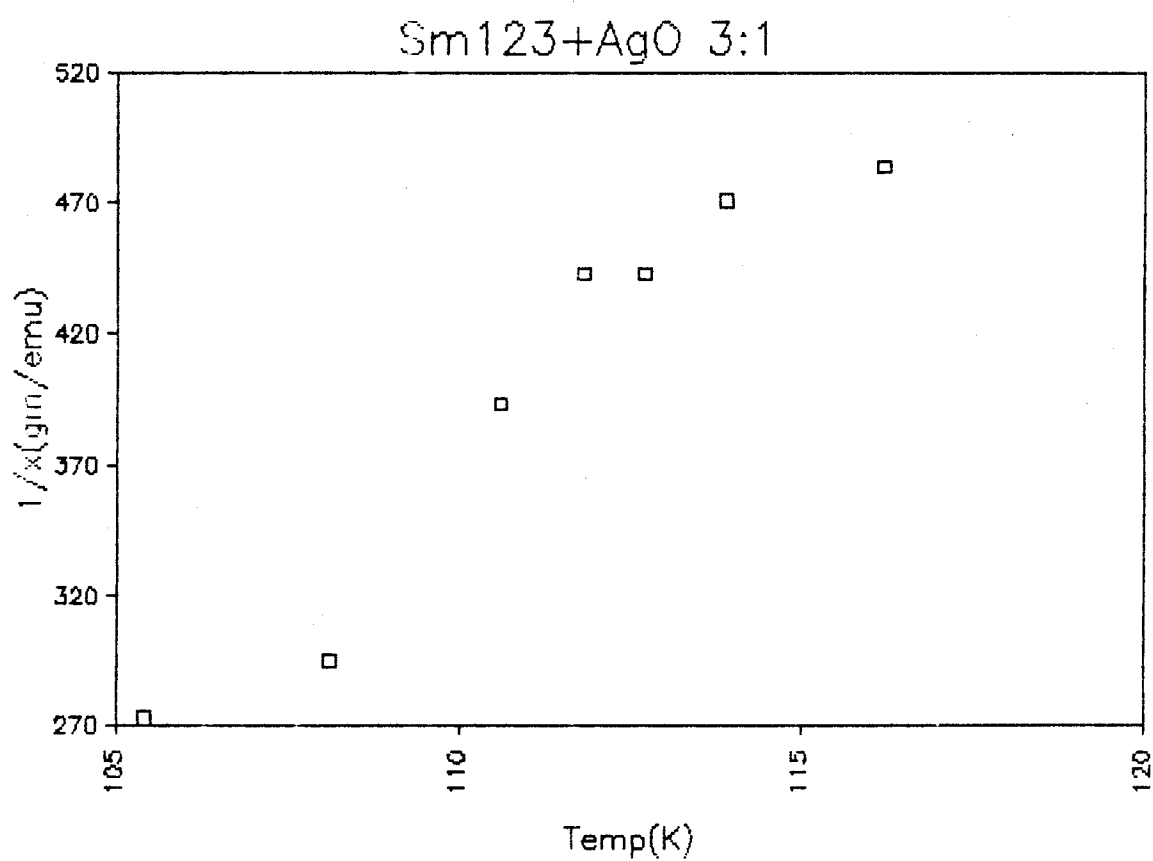


Figure 9. $1/X_m$ vs T for $T > T_c$

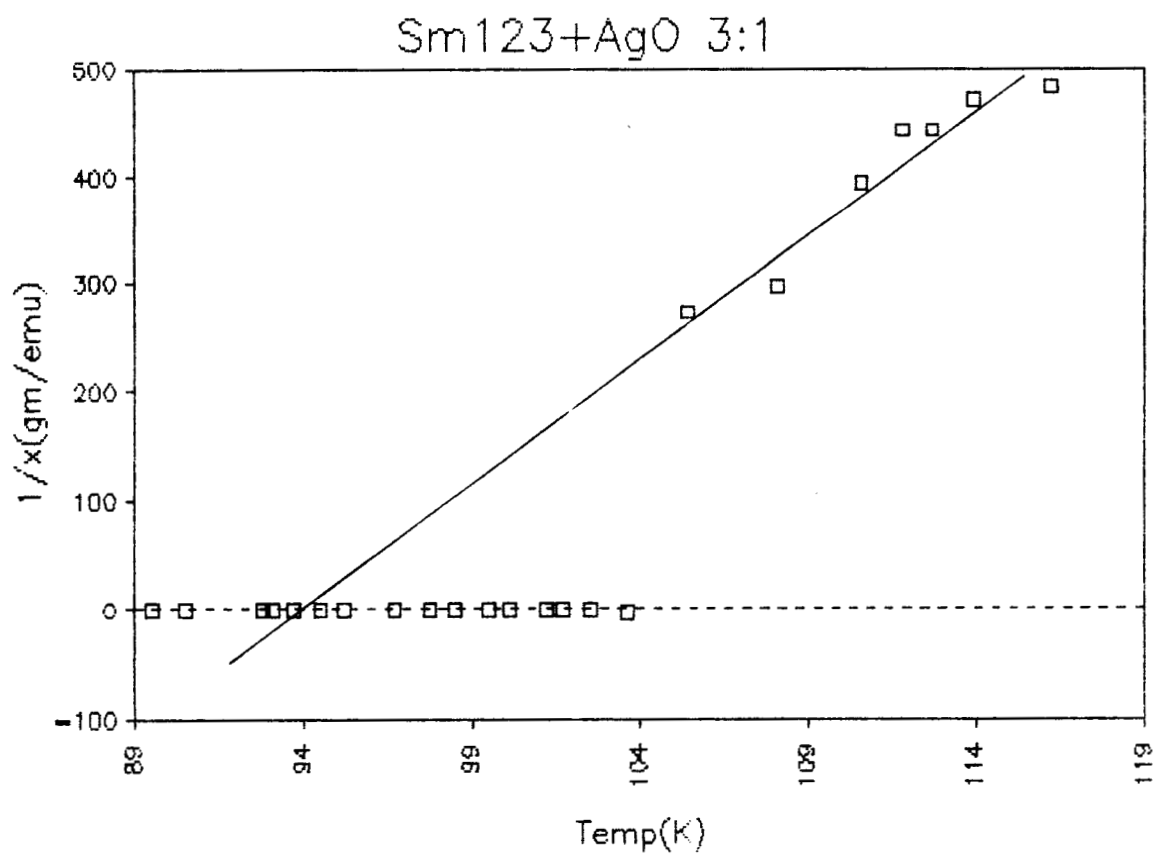


Figure 10. $1/X_m$ vs T

clearly shown in figure 7. The susceptibility begins to go diamagnetic at about 104 to 105 K. This indicates the onset of a superconducting phase. A second phase begins to show up at about 97 K and the sample becomes fully superconducting somewhere between 90 and 92 K. Resistance data on this sample, however, indicate an onset temperature of 97 K and a zero-resistance temperature of 90 K. It is evident that the resistance measurement has missed the initial phase because no complete conducting path, of this phase, existed across the sample. The second, lower temperature phase, did have such a path and it clearly shows up in the resistance data. The seven points having positive magnetization are shown in figure 8 plotted on an expanded scale. The susceptibility of a paramagnetic material obeys the Curie-Weiss law

$$X_m = C / (T - \theta) \quad (7)$$

where C is a constant, T is the absolute temperature and θ is the Curie temperature. A plot of $1/X_m$ vs T should yield a straight line of slope $1/C$ and intercept θ/C at $1/X_m = 0$. Figure 9 shows the plot of $1/X_m$ vs T for the points of positive magnetization. The least squares linear fit to these points yields the equation

$$1/X_m = -2102 + 22.5 T \quad (8)$$

which gives $\theta = 93.4$ K. Several recent papers on the magnetic properties of these compounds, using Er, Dy, and Ho as the rare earth ions, without the AgO dopant have reported values of θ which vary from -4 to -7 K^{9,10} indicating that they are antiferromagnetic. Although the samples under study here appear to obey the Curie-Weiss law, they are not antiferromagnetic. Figure 10 shows $1/X_m$ vs T for the complete data set. The intercept at 93.4 K is shown.

CONCLUSIONS AND RECOMMENDATIONS

The alternating gradient magnetometer has proved to be a powerful and versatile tool for the study of the magnetic properties of small samples. The data shown in this report represents one of the ten different compounds which were studied. These compounds were all doped with AgO and all showed an elevated Curie temperature. Furthermore, it has been demonstrated that superconducting phases can be detected which are not seen in electric resistance studies.

Three recommendations which will increase the range and usefulness of this instrument are:

1. A dewar needs to be built which will provide better temperature control during the course of the measurements.
2. A larger D.C. magnet is needed to extend the range of measurements up to several kilogauss.
3. Isolation from outside vibrations needs to be improved. This should extend the lower limit of measurement to 10^{-4} emu/gm or less.

REFERENCES

1. P. J. Flanders, J.Appl.Phys.,63,3940 (1988).
2. D. Wong, A. K. Stamper, D. D. Stancil and T. E. Schlesinger, Appl.Phys.Lett.,53,240 (1988).
3. M. K. Wu, J. R. Ashburn, C. J. Torng, P. H. Hor, R. L. Meng, L. Gao, Z. J. Huang, Y. Q. Wang and C. W. Chu, Phys.Rev.Lett.,58,908 (1987).
4. P. N. Peters, R. C. Sisk, E. W. Urban, C. Y. Huang and M. K. Wu, Appl.Phys.Lett.,52,2066 (1988).
5. M. K. Wu, J. R. Ashburn, C. A. Higgins, C. W. Fellows, B. H. Loo, D. H. Burns, A. Ibrahim, T. D. Rollin, P. N. Peters, R. C. Sisk and C. Y. Huang, Appl.Phys.Lett.,52,1915 (1988).
6. Penwalt Corporation, 900 First Ave., King of Prussia, PA 19406.
7. Handbook of Chemistry and Physics, 52nd Edition, Chemical Rubber Co., Cleveland, OH 44128
8. R. C. Sisk, private communication.
9. C. V. Tomy, S. K. Malik, R. Prasad, N. C. Soni, A. Mohan, and C. K. Gupta, J. Phys.C.,21, 3783 (1988).
10. Y. Xu, and W. Guan, Appl.Phys.Lett.,53,334 (1988).

ORIGINAL PAGE IS
OF POOR QUALITY

APPENDIX

Thermocouple Correction Table

T (C)	Emf (Tab)	Emf (Corr)	dE	<dE/dT>	<dT/dE>
0	0	.006			
-10	-.383	-.377	.3826	.0378	26.445
-20	-.757	-.750	.3736	.0369	27.127
-30	-1.121	-1.114	.3636	.0359	27.883
-40	-1.475	-1.468	.3536	.0349	28.683
-50	-1.819	-1.811	.3436	.0338	29.573
-60	-2.152	-2.144	.3326	.0328	30.521
-70	-2.475	-2.467	.3226	.0318	31.482
-80	-2.788	-2.779	.3126	.0307	32.611
-90	-3.089	-3.080	.3006	.0295	33.939
-100	-3.378	-3.368	.2886	.0283	35.318
-110	-3.656	-3.646	.2776	.0272	36.745
-120	-3.923	-3.913	.2666	.0260	38.440
-130	-4.177	-4.166	.2536	.0248	40.381
-140	-4.419	-4.408	.2416	.0235	42.527
-150	-4.648	-4.637	.2286	.0223	44.915
-160	-4.865	-4.853	.2166	.0210	47.587
-170	-5.069	-5.057	.2036	.0198	50.596
-180	-5.261	-5.249	.1916	.0185	54.159
-190	-5.439	-5.426	.1776	.0171	58.602
-200	-5.603	-5.590	.1636	.0164	61.109

$$T = T_0 + \langle dT/dE \rangle (E - E_0)$$

N89-21740

1988

**NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA**

GUIDANCE FOR AN AEROASSISTED ORBITAL TRANSFER VEHICLE

Prepared by:	Kenneth R. Hall
Academic Rank:	Associate Professor
University and Department:	Mississippi State University Aerospace Engineering

NASA/MFSC:

Laboratory:	Engineering
Division:	Guidance and Control
Branch:	Flight Mechanics

MSFC Colleague:	Roger Burrows
Date:	August 24, 1988
Contract No:	NGT 01-002-099 The University of Alabama

Guidance For An Aeroassisted Orbital Transfer Vehicle

by

Kenneth R. Hall

Associate Professor of Aerospace Engineering

Mississippi State University

Mississippi State, MS 39762

ABSTRACT

The use of atmospheric drag for slowing satellites in high energy, high apogee orbits to a lower energy, lower apogee orbit about the earth is investigated. The high energy orbit is assumed to intercept the earth's atmosphere. Guidance for the atmospheric phase of the trajectory may be done using the aerodynamic forces generated by the passage through the atmosphere. This research was concerned with the investigation of several methods of guidance during the atmospheric phase to cause a significant reduction in the final velocity as the vehicle leaves the atmosphere. In addition, the velocity direction was controlled to exit to a desired target orbit. Lastly excess aerodynamic lift was used to effect a plane change between the entry orbit plane and the exit orbit plane to achieve a desired orbit plane.

The guidance methods were applied to a 3 degrees-of-freedom simulation which included an oblate earth gravity model and a rotating atmosphere. Simulation results were compared on the basis of speed of computation of the guidance parameters and amount of added velocity necessary to achieve the desired orbit.

ACKNOWLEDGEMENTS

I am grateful for the opportunity to participate in the Summer Faculty Fellowship Program at Marshall Space Flight Center. I am especially grateful for the support and encouragement of my NASA colleague, Roger Burrows, and the members of his branch who supplied help cheerfully, whenever I asked. I am also deeply indebted to Wayne Deaton, head of the Flight Mechanics Branch, and Wilton Causey, deputy division chief, for their aid. Lastly but not least, Greg Dukeman worked on a similar task and enjoyed some of the same headaches and successes.

I would like to thank Dr. Mike Freeman for his encouragement and guidance throughout the program. The experience at Marshall Space Flight Center was very enjoyable. I am indebted to the National Aeronautics and Space Administration and to the American Society of Engineering Education for providing me with this unique opportunity.

Guidance For An Aeroassisted Orbital Transfer Vehicle

There is currently underway a study concerning the use of atmospheric braking for vehicle transfer from a high energy orbit to a lower energy orbit about the earth. The use of atmospheric drag is seen as an excellent way to decrease the vehicle speed in the high energy orbit to the speed required for the lower energy orbit. In addition to the decrease in the speed of the vehicle, the plane of the new orbit may not coincide with the plane of the original orbit. This study is concerned with the investigation of the guidance requirements for the atmospheric braking portion of the vehicle trajectory with targeting to a final orbit.

Problems of this type are generally analyzed by the iterative methods of perturbation analysis. The trouble with this case is that the time required for the calculation of the solution to the two point boundary value problem arising from the perturbation analysis may be too time-consuming for use in the vehicle guidance system. The atmospheric flight phase of the trajectory is characterized by sensitivity to the time in which maneuvers are accomplished. This sensitivity may even lead to atmospheric capture of the vehicle or to a final orbit which is nowhere near the target orbit. In addition to the time sensitivity, the atmospheric flight is characterized by large changes in atmospheric density and dynamic pressure on the vehicle. Due to the requirement that the guidance system operate the guidance loop rapidly and effectively, a reduced complexity set of differential equations was chosen for the guidance system model and approximate solutions were employed for guidance and targeting. These solutions were applied to a 3 degree of freedom, point mass simulation with full complexity simulation differential equations including an atmospheric model with perturbations which rotates with the earth and an oblate earth model with the J_2 harmonic terms included in the gravity model.

The figure below represents a shuttle launched experiment for testing the vehicle and the guidance system. The shuttle launches the vehicle into a high apogee lob trajectory. Just before returning to the original launch point, the on-board engine increases the speed and changes the velocity direction to simulate a vehicle returning from a geosynchronous orbit. The new orbit intersects the "reasonable" atmosphere of the earth at an altitude of 400000 feet, passes through the atmospheric transfer portion of the trajectory during which atmospheric forces are employed to

effect velocity and orbital plane changes to target to the final desired orbit at the exit from the atmosphere. The target orbit is characterized by an apogee altitude of 200 miles and a specific orbital plane inclination. At the apogee of the achieved final orbit, additional changes are made in the orbit to circularize the final orbit and to reach the desired orbital inclination. For this simulation, the specified entry conditions and the desired exit conditions are tabulated below.

ENTRY CONDITIONS

H=400000 FT

V=33828 FPS

$\gamma = -4.5$ DEG

i=27. DEG

EXIT CONDITIONS

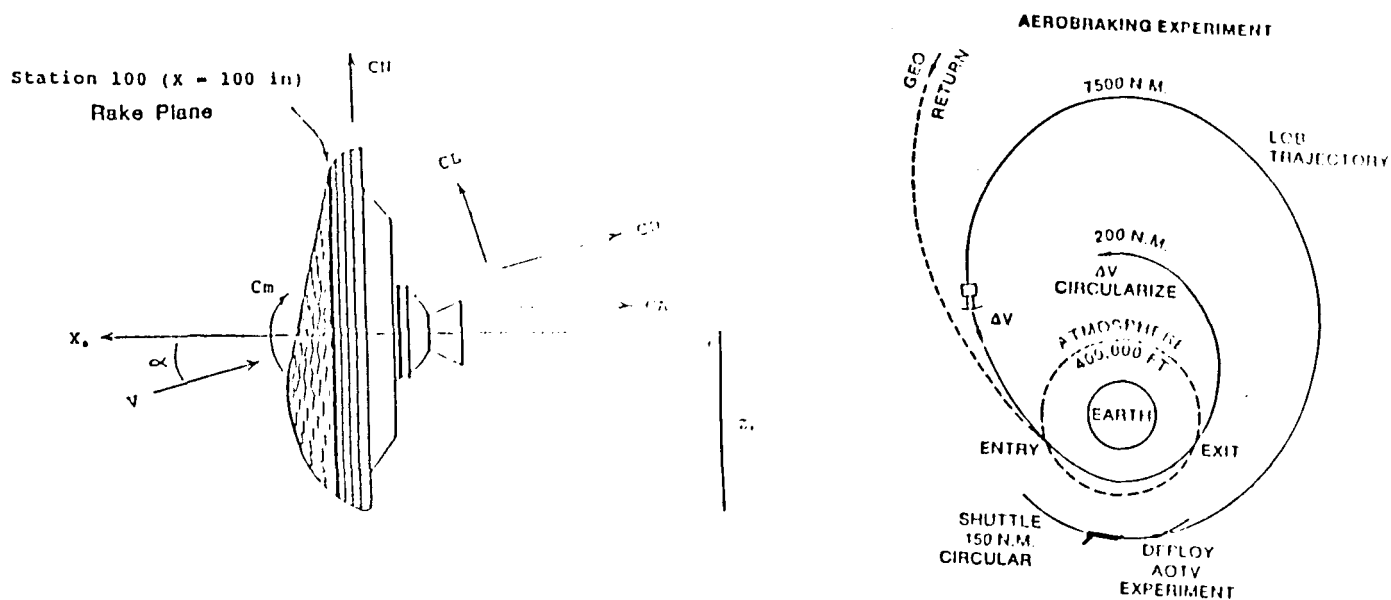
H=400000 FT

HP=50. MILES

HA=200. MILES

i=28.5 DEG

FIGURE 1. PROBLEM DEFINITION IN GRAPHICAL FORM



During the atmospheric transfer portion of the trajectory, the AOTV is assumed to maintain a constant angle of attack with the resulting constant lift coefficient (C_L) and constant drag coefficient (C_D). Control of the AOTV during this phase of the trajectory is done by rolling the vehicle to orient the lift force to cause in-plane vertical acceleration for altitude control and out-of-plane transverse acceleration for control of the orbital plane inclination. The equations of motion used in the guidance model may be written as follows:

GUIDANCE EQUATIONS OF MOTION

$$\frac{dV}{dt} = - \frac{C_D S}{2M} \rho V_R^2 - \frac{G}{V} \frac{dH}{dt}$$

$$\frac{d^2 H}{dt^2} = \frac{V^2}{R} - G + \frac{C_L S}{2M} \rho V_R^2 \cos(\phi)$$

In the equations above, V is the inertial velocity magnitude, V_R is the airspeed or speed relative to the rotating atmosphere, ρ is the density of the atmosphere at altitude H above the surface of the earth, G is the gravitational acceleration at altitude H above the surface, R is the distance from the center of the earth to the AOTV, M is the constant mass of the AOTV, C_L and C_D are the lift and drag coefficients of the AOTV, and ϕ is the roll angle used to control the vehicle. Using the second equation, the guidance equation may be specified as follows:

GUIDANCE ROLL COMMAND EQUATION

$$\cos(\phi)_{\text{COM}} = \frac{2M}{C_L S \rho V_R^2} \left[\left(\frac{d^2 H}{dt^2} \right)_{\text{COM}} + G - \frac{V^2}{R} \right]$$

Two items are apparent from the equation above: (1) The roll angle is specified in cosine form, and (2) The roll angle cosine depends upon the specification of the vertical acceleration. For the first item, use the guidance roll command equation to specify the cosine of the roll angle. Use the

direction of the desired plane change maneuver to determine the sign of the sine of the roll angle. The combination of the roll angle sine and cosine will then uniquely determine the roll angle. A characteristic of this method is that the roll angle "ratchets" in that the amount of roll angle for plane control may be too great and the sign of the sine may need to be reversed periodically to stay within an allowable deadband of the desired final inclination angle. Typical trajectories will display four or more sign reversals on the sine term during the atmospheric pass. The second item, the commanded vertical acceleration in H is a different story. This study explored four methods of completing the roll command equation.

METHOD 1. ASSUME A CONSTANT ALTITUDE RATE

For a constant altitude rate, the commanded acceleration is zero which simplifies the guidance equation. For the entry phase, this application is straightforward as the initial conditions to start the "equalibrium glide" are specified. For the exit phase, the initial conditions are not specified and one must use iteration to determine the speed with which to begin the exit phase and the altitude rate command for the exit phase, which in combination yield the desired exit conditions. (See C. Cerimele's paper listed in the references for a detailed explanation of this method.) This method does succeed but relies upon the iteration technique for commands during the exit phase. In actual practice, the commanded roll angle is augmented by commands which will produce desired values of altitude rate and dynamic pressure, as otherwise the command would drive the AOTV down the constant altitude rate path into the lower atmosphere from which it would not be able to initiate an exit phase.

CONSTANT ALTITUDE RATE ROLL COMMAND

$$\cos(\phi)_{\text{COM}} = \frac{2 M}{C_L S \rho V_R^2} \left[G - \frac{V^2}{R} - G_H \left(\frac{dH}{dt} - \frac{dH}{dt}_{\text{COM}} \right) - G_Q (Q_B - Q_{B_{\text{COM}}}) \right]$$

The G_Q and G_H terms are dynamic pressure and altitude rate gains used in combination with altitude rate, $\frac{dH}{dt}$, and dynamic pressure, Q_B , to produce the desired trajectory upon entry. For the exit trajectory, G_Q is zero and the $\frac{dH}{dt}_{\text{COM}}$ is determined iteratively.

METHOD 2. ASSUME A CONSTANT ROLL ANGLE COSINE

This is another easily implemented option as the constant roll angle cosine eliminates any control determination for the vertical acceleration. The sign of the roll angle sine is still chosen to produce the desired plane change. The determination of the actual roll angle cosine is done in an iterative manner using a numerical perturbation in the constant roll angle to determine numerical partials of the target conditions with respect to the chosen roll angle cosine. Again, this method will work, but it depends upon solving a two point boundary value problem iteratively, and speed of guidance calculations may be its demise.

CONSTANT ROLL ANGLE COSINE COMMAND

$$\cos(\phi)_{\text{COM}} = \text{CONSTANT.}$$

METHOD 3. ASSUME AN ALTITUDE TIME HISTORY

For this method, assume a polynomial in time for the altitude. Include in the polynomial enough coefficients to be determined to allow the method to fit an optimum criteria. For example:

ASSUMED ALTITUDE POLYNOMIAL

$$1. \text{ ASSUME } H = H_1 + \frac{dH_1}{dt} T + \frac{d^2H_1}{dt^2} \frac{T^2}{2} + \frac{d^3H_1}{dt^3} \frac{T^3}{6}$$

$$\frac{dH}{dt} = \frac{dH_1}{dt} + \frac{d^2H_1}{dt^2} T + \frac{d^3H_1}{dt^3} \frac{T^2}{2}$$

$$\frac{d^3H}{dt^3} = \frac{2}{T_2^2} \left[\frac{dH_2}{dt} - \frac{dH_1}{dt} - \frac{d^2H_1}{dt^2} T_2 \right]$$

$$\frac{T_2^2}{6} \left[\frac{1}{6} \frac{d^2H_1}{dt^2} \right] + T_2 \left[\frac{2}{3} \frac{dH_1}{dt} + \frac{1}{3} \frac{dH_2}{dt} \right] + \left[H_1 - H_2 \right] = 0$$

$$T_2 = \frac{-B \pm [B^2 - 4AC]^{1/2}}{2A}$$

GUESS UNKNOWN VALUES AND ITERATE TO A SOLUTION.

$$\frac{d^2 H}{dt^2} \text{ COMMAND} = \frac{d^2 H_1}{dt^2} + \frac{d^3 H}{dt^3} T$$

$$\cos(\phi) \text{ COMMAND} = \frac{2 M}{C_L S \rho V_R^2} \left[\left(\frac{d^2 H}{dt^2} \right) \text{ COMMAND} + G - \frac{V^2}{R} \right]$$

The method above has merit in that it will allow targeting to a specific altitude with time, altitude rate, and vertical acceleration involved in the iteration. Analytical approximations allow quadrature approximations for integration which allow speed increases in iterative passes. It is still an iterative scheme but is a fast iterative scheme with explicit targeting.

METHOD 4. ASSUME A VERTICAL ACCELERATION FUNCTION

This last method involves assuming that the AOTV will act like a damped harmonic oscillator with specified damping ratio and natural frequency. The vertical acceleration command is then easily determined from the following equation:

ASSUMED OSCILLATOR FORM

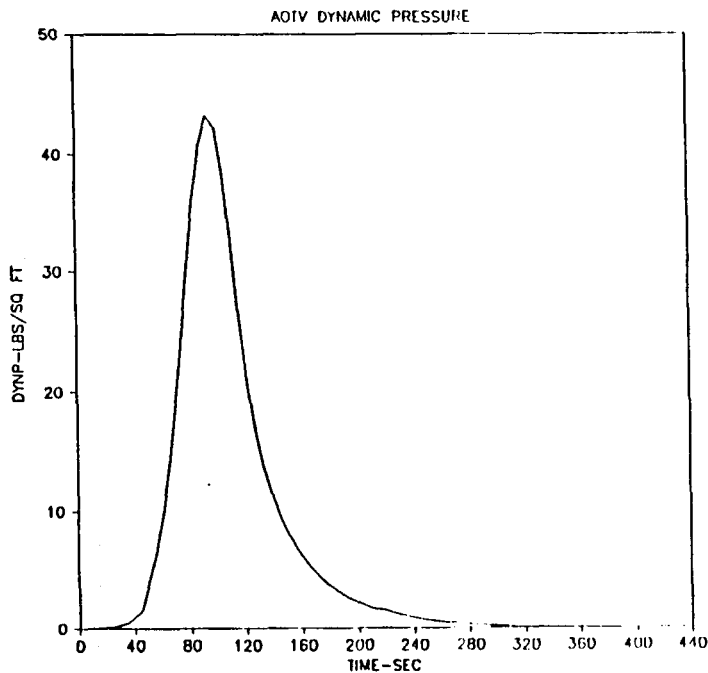
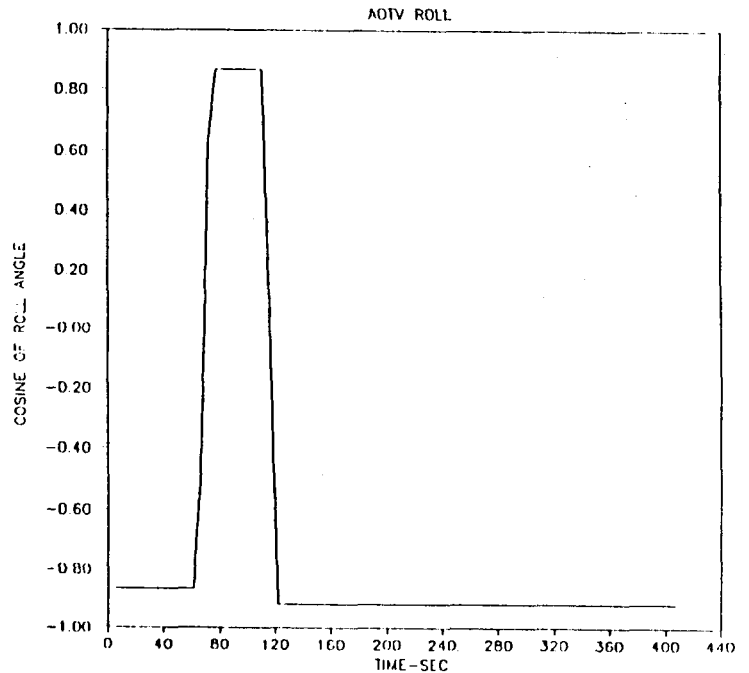
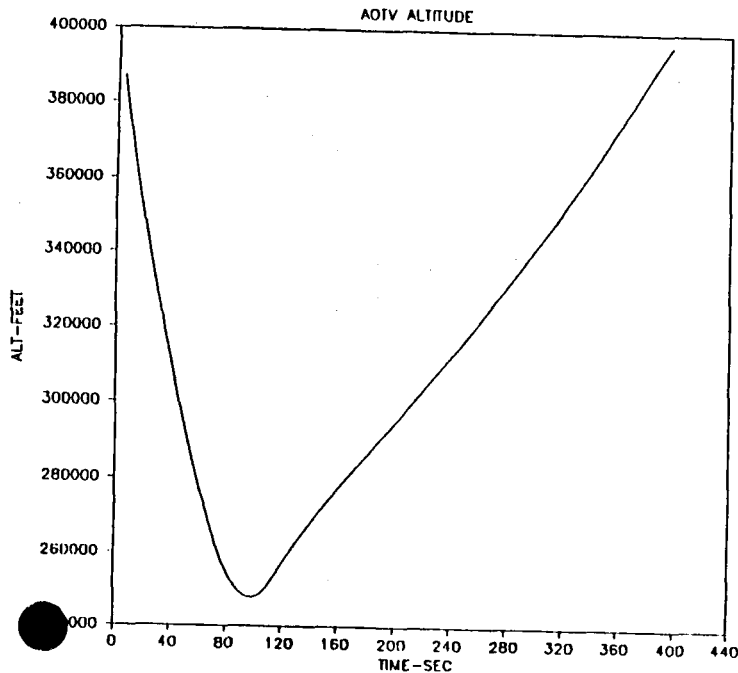
$$\frac{d^2 H}{dt^2} \text{ COM} = -2 \zeta \omega_N \frac{d H}{dt} - \omega_N^2 (H - H_{\text{COM}})$$

In the oscillator equation above, the selection of damping ratio, ζ , and natural frequency, ω_N , are equivalent to specifying the manner in which the AOTV system will react. The appearance of H_{COM} in the equation allows targeting to a specific altitude on the entry phase and assures that the altitude rate will be nulled at that altitude. For the exit phase, this same form can be used with different system characteristics.

$$\cos(\phi) \text{ COM} = \frac{2 M}{C_L S \rho V_R^2} \left[\left(\frac{d^2 H}{dt^2} \right) \text{ COM} + G - \frac{V^2}{R} \right]$$

The outstanding benefits of the guidance scheme listed above is that it is fast (with no iteration), it allows altitude targeting, and the same form of guidance can be used in the entry phase and the exit phase with appropriate changes in the system characteristics.

The figures shown below represent the trajectory shape for the atmospheric phase, with the attendant dynamic pressure, and a roll cosine profile. The roll cosine does not exhibit the "ratcheting" which the roll angle does as the "ratcheting" is due to sign reversals on the sine function.



LIST OF REFERENCES

1. Cerimele, C.J. and Gamble, J.D.: "A Simplified Guidance Algorithm for Lifting Aeroassist Orbital Transfer Vehicles": AIAA 23rd Aerospace Sciences Meeting, AIAA 85-0348., Jan., 1985
2. Higgins, J.P.: "An Aerobraking Guidance Concept For A Low L/D AOTV": CSDL OTV memo 10E-84-04. May, 1984
3. Loh, W.H.T.: Dynamics and Thermodynamics of Planetary Entry: Prentice-Hall, 1963
4. Duncan, R.C.: Dynamics of Atmospheric Entry: McGraw Hill, 1962
5. Bate, R.R., Mueller, D.D., White, J.W.: Fundamentals of Astrodynamics: Dover, 1971
6. Bryson, A.E., Jr., and Ho, Y.C. : Applied Optimal Control: Hemisphere Publishing Company, 1975

N89-21741

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

TRANSIENT NATURAL CONVECTION HEAT AND MASS TRANSFER
IN CRYSTAL GROWTH

Prepared by:	Samuel S. Han
Academic Rank:	Associate Professor
University and Department:	Tennessee Technological University Mechanical Engineering
NASA/MSFC:	
Laboratory:	Structure and Dynamics
Division:	Earth Science and Applications
Branch:	Fluid Dynamics
MSFC Colleague:	Charles F. Schafer Cheryl Morroni
Date:	August 10, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

TRANSIENT NATURAL CONVECTION HEAT AND MASS TRANSFER IN CRYSTAL GROWTH

by

Samuel S. Han
Associate Professor of Mechanical Engineering
Tennessee Technological University
Cookeville, Tennessee 38505

ABSTRACT

A numerical analysis of transient combined heat and mass transfer across a rectangular cavity is performed by a numerical method based on the SIMPLE algorithm. The physical parameters are selected to represent a range of possible crystal growth in solutions.

Numerical results are compared with available experimental data to confirm the accuracy of the results. Good quantitative agreements are obtained for the average mass transfer rate across the cavity. Also, qualitative agreements are observed for the global development of thermal and solute fields.

It is found that the thermal and solute fields become highly oscillatory when the thermal and solute Grashof numbers are large. Oscillations are probably caused by number of different instability mechanisms. By reducing the gravity some of these instabilities were made to disappear at the lower Grashof numbers.

Transient temperature and solute distribution near the crystal growing surface are highly non-uniform at the higher Grashof numbers. These non-uniformities are less severe in the reduced gravity environments but still exist. The effects of convection on the rate of average mass transfer are more than one order of magnitude higher than those of conduction in the range of Grashof numbers studied. Dependency of mass transfer rate on the Grashof number indicates that the convection effects may not become negligible even in the microgravity environments for the range of parameters investigated.

ACKNOWLEDGEMENT

I would like to offer my deepest gratitude to Dr. Charles F. Schafer and Ms. Cheryl Morroni, my MSFC colleagues, for giving me an opportunity to learn the importance of single crystals and to appreciate intricate transport processes involved in crystal growth.

Without help, advice and guidance from many colleagues in the Fluid Dynamics Branch, my stay at Marshall could be a difficult one. Many thanks to Drs. Chen, Ramachandran, Decker and Yeshola and to Mr. Trinh and Ms. Janti.

Dr. Michael Freeman and Ms. Ernestine Cothran, Co-directors of 1988 program, deserve a special gratitude for their excellent administrative help and guidance.

ORIGINAL PAGE IS
OF POOR QUALITY

1. INTRODUCTION

Single crystals are important parts of modern scientific and engineering equipments because of their unique electro-magnetic properties. Most single crystals are made by one of three methods: vapor growth, solution growth and melt growth.

The final transport process in crystal growth is the phase change from liquid to solid at the growing interface. Mass transport is dominantly controlled by the mass diffusion near the interface between the liquid and solid. Thermodynamic and transport properties near the growing surface are however intimately connected to the thermal and solute fields far away from the interface. These far-field behaviors are determined by the convective heat, mass and momentum transport processes [1].

Convective transport processes tend to expedite the mass transport of crystal materials across the ampoule. Highly non-uniform temperature and solute distributions inherent in convection, however, have undesirable effects on the quality of crystals. Convection effects, in parts, are believed to be responsible for non-uniform properties of single crystals grown in vapor and solution growth and helical segregation of impurities in crystals grown in melt growth [2,3].

To improve the quality of single crystals, quantitative assessment of convection effects are required. In recent years, many research efforts, both analytical and experimental approaches, are reported in literatures [4,5,6,7]. Since the actual processes involved in crystal growth are quite complex and some of the physical processes are not well understood, most of these investigations focused on a few limited aspects of convective processes.

The objective of the present research is to investigate the transient heat and mass transfer in a two-dimensional rectangular ampoule in a solution growth by a numerical analysis. Methods of analysis are presented in section 2, numerical results are presented and discussed in section 3, followed by conclusions and recommendations in section 4.

II. METHODS OF ANALYSIS

REF ID: A61111
PAGE 18
OF FOUR QUALITY

Description of the Problem

Actual physical mechanisms involved in crystal growth are time-dependent, three-dimensional convection-diffusion processes with a phase change at the crystal growing surface. Enthalpy of phase transformation might be involved in certain applications. In general, crystal growing surfaces are not flat and they move at the rate of growing crystals. Boundary conditions imposed on the ampoule surfaces can be quite complex. Physical properties of the mixture in the ampoule might be strongly dependent upon the solute concentration and temperature.

The present research is concerned with the prediction of solute and temperature fields encountered in solution growth. To make the analysis feasible, number of simplifying assumptions are introduced in formulating the problem. The geometry of the ampoule is rectangle with a fixed aspect ratio. The left vertical wall is kept at a higher constant temperature and a higher concentration of crystal material while the right vertical wall is kept at a lower temperature and a lower concentration of crystal material. A binary mixture made of a nutrient crystal material and a carrier fluid fills the cavity of the ampoule. The crystal materials are dissolved from the left wall and deposited on the right wall. Horizontal surfaces are adiabatic and impermeable to both components. The vertical walls are also impermeable to the carrier fluid. Figure 1 shows the geometry of the ampoule considered in the present analysis.

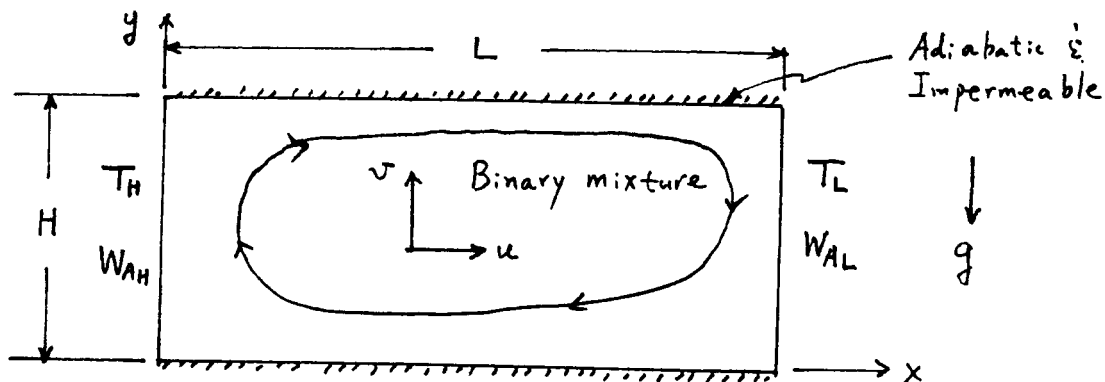


Figure 1. Geometry of the problem

Governing Equations

Assuming a two-dimensional incompressible flow of Newtonian fluid for the mixture, conservation of mass, momentum, energy and crystal material (hereafter called species A) may be expressed by the following partial differential equation:

$$\frac{\partial \rho \phi}{\partial t} + \frac{\partial}{\partial x} [\rho u \phi - \Gamma \frac{\partial \phi}{\partial x}] + \frac{\partial}{\partial y} [\rho v \phi - \Gamma \frac{\partial \phi}{\partial y}] = S \quad (1)$$

where ϕ is the dependent variable, Γ is the diffusion coefficient and S is the source. These values are listed in Table 1 for each equations. In this expression, t is time, x and y are horizontal and vertical distance, u and v are fluid velocity in x - and y -directions, E is the internal energy, W_A is the mass fraction of species A, ρ is the density of the mixture, T is the temperature, p is the pressure, μ is the dynamic viscosity, k is the thermal conductivity and D_{AB} is the mass diffusivity.

Table 1. Components of governing equations

equation	ϕ	Γ	S
mass	1	0	0
x-momentum	u	μ	$-\frac{\partial p}{\partial x}$
y-momentum	v	μ	$-\frac{\partial p}{\partial y} + \rho g \beta (T - T^0) + \alpha_A \rho g (W_A - W_A^0)$
energy	E	k/c_v	0
concentration	W_A	ρD_{AB}	0

It is assumed that the density of the fluid remains at a constant value and the effects of heat and mass transfer from the vertical walls appear as the buoyancy force along the y -direction. This is known as a Boussinesq approximation and is valid when the temperature difference between the walls are small and the mass fraction of species A is small (dilute solution). In terms of Boussinesq approximation, the density of the mixture is given by

$$\rho = \rho^0 [1 - \alpha_A (W_A - W_A^0) - \beta (T - T^0)] \quad (2)$$

where, β and α_A are the thermal and solute expansion coefficient, respectively. They are

$$\beta = -\frac{1}{\rho} \left(\frac{\partial \rho}{\partial T} \right)_p \quad (3)$$

and

$$\alpha_A = M_0 (M_B - M_A) / M_A M_B \quad (4)$$

where

$$1/M_0 = W_A^0/M_A + W_B^0/M_B \quad (5)$$

M_A and M_B are the molecular weight of species A and B (carrier fluid), respectively. Superscript '0' implies a reference state.

Initial and Boundary Conditions

Initially, a mixture with a constant temperature (T^0) and mass fraction (W_A^0) is in hydrostatic equilibrium under a given gravity (g). At $t > 0$, temperature at the left wall is increased by a small amount and the temperature at the right wall is decreased by the same amount. Heat transfer occurs from the left wall to the adjacent fluid by conduction. Density of heated fluid, in general, becomes small and the thermal buoyancy force acts in the opposite direction to the gravity. Cooled fluid adjacent to the right wall moves in the same direction as the gravity. This buoyancy force due to the thermal effects drive the convection in the ampoule. When the concentration of species A at the left wall is increased from its initial state and decreased by the same amount at the right wall, convection occurs due to solutal buoyancy force. If $M_A > M_B$, buoyancy force due to mass transfer will act downward near the left wall and upward near the right wall. In this case, two buoyancy forces will act in the opposite direction (opposing convection). If $M_A < M_B$, then two buoyancy forces will act in the same direction (aiding convection).

Mathematically, the initial conditions are

$$\begin{aligned} u(x, y, t=0) &= v(x, y, t=0) = 0 ; \\ \rho(x, y, t=0) &= \rho^0 ; \quad T(x, y, t=0) = T^0 \\ W_A(x, y, t=0) &= W_A^0 \end{aligned} \quad (6)$$

and the boundary conditions are

$$\begin{aligned} T(x=0, y) &= T_H ; \quad T(x=L, y) = T_L ; \\ W_A(x=0, y) &= W_{AH} ; \quad W_A(x=L, y) = W_{AL} ; \\ \left. \frac{\partial T}{\partial y} \right|_{x, y=0} &= \left. \frac{\partial T}{\partial y} \right|_{x, y=H} = \left. \frac{\partial W_A}{\partial y} \right|_{x, y=0} = \left. \frac{\partial W_A}{\partial y} \right|_{x, y=H} = 0 ; \end{aligned} \quad (7)$$

$$\begin{aligned} u(x, y=0) &= v(x, y=0) = u(x, y=H) = v(x, y=H) = 0 ; \\ u(0, y) &= v(0, y) = u(L, y) = v(L, y) = 0 . \end{aligned}$$

Nondimensional Parameters

Selecting H , ν/H , ΔT and ΔW_A as the characteristic values, it can be readily shown from the governing differential equations that there are five nondimensional parameters which characterize the transport processes in the ampoule. They are

$$\begin{aligned} A &= H/L ; \quad Pr = \nu/\alpha ; \quad Sc = \nu/D_{AB} ; \\ Gr_T &= g\beta\Delta TH^3/\nu^2 ; \quad Gr_S = g\alpha_A\Delta W_A H^3/\nu^2 \end{aligned} \quad (8)$$

where $\Delta T = T_H - T_L$ and $\Delta W_A = W_{AH} - W_{AL}$.

Numerical Method

The numerical method used to solve the given equations subject to the initial and boundary conditions is a modified version of the SIMPLE algorithm [8]. The SIMPLE and its many variants have been successfully applied to a wide range of fluid flow and heat transfer problems. The present method follows exactly the same procedures as described in ref.[8] except a simple modification to account for the density change due to the pressure change [9]. In solving the continuity equation, a recommendation given in ref.[10] is used to speed up the convergence.

III. RESULTS AND DISCUSSIONS

Physical Parameters

The physical size of the ampoule and the mixture properties are selected to match with some of experimental data reported in ref.[7] in terms of nondimensional parameters. Three nondimensional parameters are fixed at constant values. They are $A=0.498$, $Pr=7.07$ and $Sc=2140$. Remaining two nondimensional parameters are varied to analyse the effects of their magnitude and the effects of opposing and aiding buoyancy forces. Table 2 shows these parameters for 5 cases studied in the present investigation.

Table 2. Grashof numbers (Gr_T, Gr_S) and buoyancy ratio (N)

Case	Gr_T	Gr_S	$N=Gr_S / Gr_T$
1	4.36E+5	8.12E+6	18.62
2	4.36E+5	-8.12E+6	-18.62
3	4.36E+4	8.12E+5	18.62
4	4.36E+4	-8.12E+5	-18.62
5	4.36E+3	8.12E+4	18.62

Experimental data are available for cases 1 and 2. They are used to check the accuracy of the present analysis. Smaller Grashof numbers used in cases 3-5 are obtained by reducing the gravity to 0.1 g and 0.01 g, where $g=9.81 \text{ m/s}^2$. This approach is used to examine the effects of the reduced gravity when the mixture properties remain constant.

Numerical Parameters

Numerical results depend on the choice of numerical

parameters, such as the mesh sizes, time steps, relaxation coefficients and a convergence criterion.

The number of control volumes and their arrangements are usually determined by a trial and error method until the numerical results agree with known data or become grid independent. It was reported in a previous investigation that 26 (horizontal) by 20 (vertical) control volumes gave sufficiently accurate average heat or mass transfer rate when the heat or mass transfer occurs separately [11]. A highly nonuniform grid arrangement was used in that study. The control volume sizes near the vertical walls were selected such that thin solute boundary layer developed on these surfaces contained more than one control volumes. In the same study, 26 by 20 grid system was also used for combined heat and mass transfer. Overall average mass transfer rate was seen in good agreement with experimental data. However, numerical results failed to exhibit intricate three layered convection loops observed in the experiment [1,7]. Subsequent experiments on the grid arrangement (30 by 28 and 40 by 40) revealed that increasing grid number in the vertical direction gave improved resolution of flow structure but the average mass transfer rates were not significantly changed beyond those obtained by the 26 by 20 system. In the present analysis, a 40 by 40 grid system is used for all cases.

In the present study, numerical integration begins with a physically realistic initial condition. Consequently, convergence of numerical solutions at each time level is achieved with a few iterations provided that a proper time step is used. At the beginning of computation an extremely small time step is used. In the subsequent computations, time steps are automatically adjusted depending on whether the number of iteration to achieve the convergence is below or above the preset minimum (2) and maximum (10) number of iterations. The convergence of flow quantities in each time step is assumed when the relative residual mass in every control volume is less than 0.1 %.

The relaxation coefficients for the momentum equations were 0.5 for both directions and no relaxation is used for the pressure and energy equations.

Transient Global Flow Fields

Figure 2 shows transient global development of momentum, thermal and solute fields for case 1 at three different time steps. The velocity vector (top panel) and stream function (second from top) distribution show the effects of mass transfer on the convection. Since the thermal diffusivity is much larger than the mass diffusivity ($Sc/Pr = 300$), the convection loop in the ampoule is determined by the thermal effects only at the beginning stage of convection. However, the presence of mass transfer soon begins to modify the thermally driven convection through the double diffusive effects [12].

ORIGINAL PAGE IS
OF POOR QUALITY

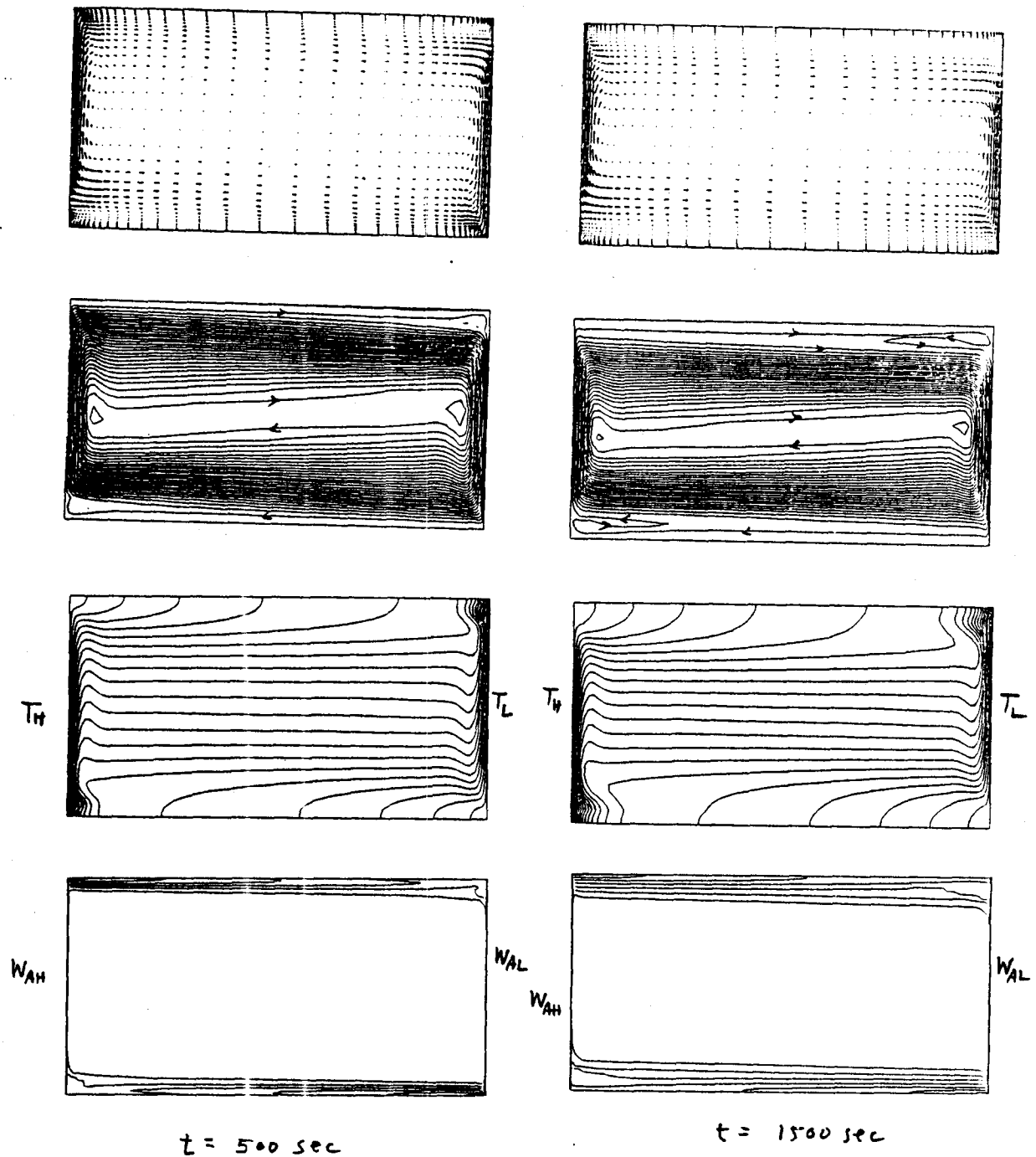


Figure 2. Transient velocity, temperature and solute fields at $t=500 \text{ sec}$, $t=1500 \text{ sec}$ and $t=4000 \text{ sec}$ for case 1 and $t=4000 \text{ sec}$ for case 2.

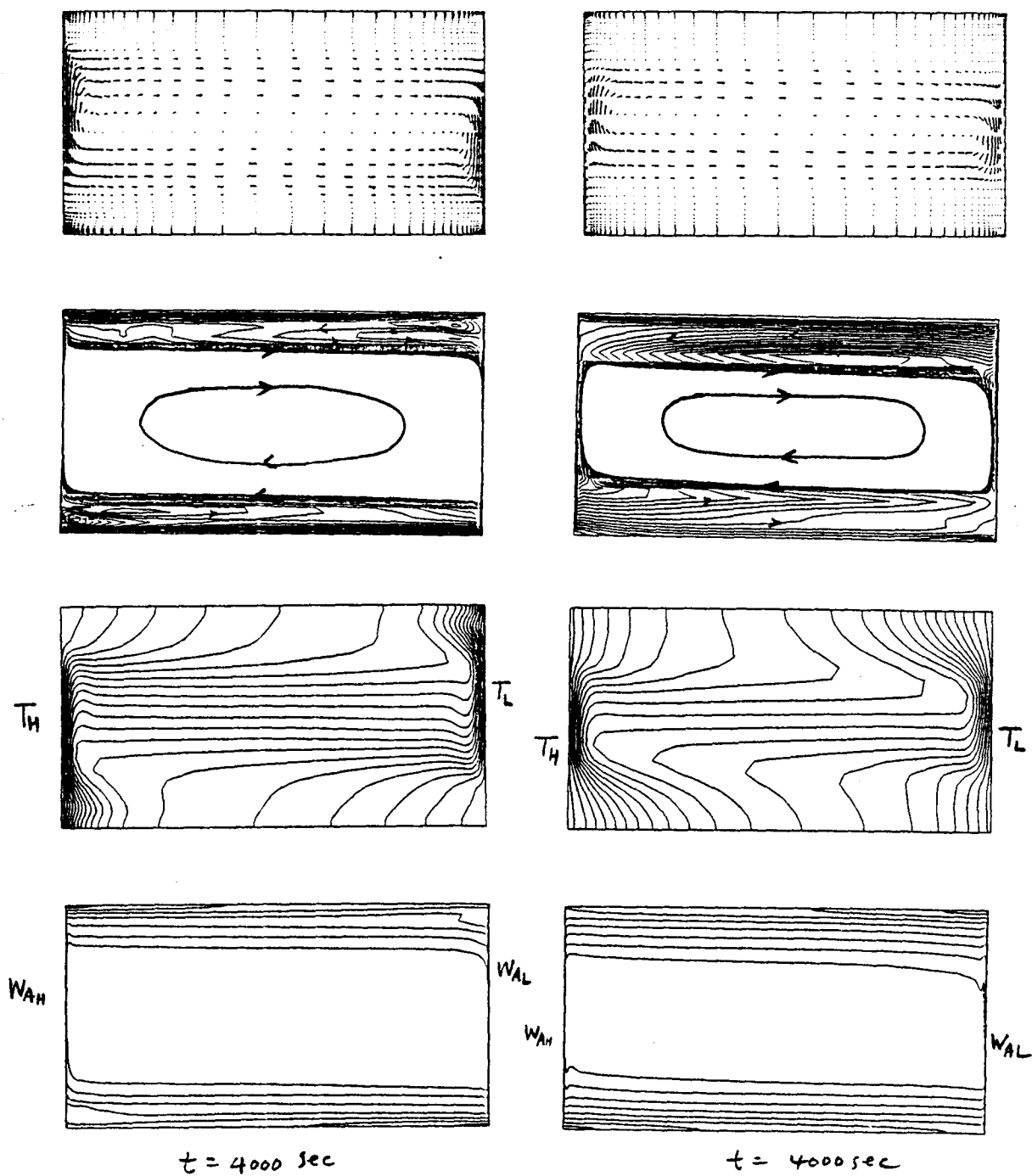


Figure 2. continued (see captions on the previous page)

The most striking feature of the flow field is the appearance of clockwise convection loops near the horizontal surfaces. This feature is in agreement with experimental data reported in ref. [1,7]. The physical explanations offered in ref. [7] are as follows. As species A diffuses from the left wall to the adjacent mixture, it is carried upward by convection and a thin solute boundary layer develops at the top surface. As the mass transfer continues, the solute boundary layer near the horizontal surface thickens. When the solute boundary layer thickness reaches certain level, some of species A in this solute boundary layer is entrained by the thermally driven convection and swept away from the solute boundary layer. At the right wall, some of this entrained portion of species A can not diffuse to the right wall fast enough to be pulled down by the thermally driven convection at the right wall since mass diffusion is much slower than thermal diffusion. This left over light fluid is then pushed toward the left wall and a clockwise convection loop at the upper right corner of the ampoule appears. Similar double diffusive effects result in a clockwise convection loop at the lower left corner. As time progresses, horizontal solute boundary layer thickness increases (bottom panel) and solute driven weak circulation loops penetrate deeper into the cavity.

Temperature field in the cavity is strongly influenced by the solute transport (third panel). Temperature inversion with a steep gradient is apparent where the solute driven convection loops exist. Therefore, there exist destabilizing temperature gradient and stabilizing solute gradient in these region and flow oscillation of over-stable mode can be triggered [13]. As expected in a high thermal Grashof number convection, the temperature in the core region of the ampoule is completely stratified.

The last column in Figure 2 shows the global flow properties for case 2 at $t=4000$ sec. The only difference between case 1 and case 2 is that buoyancy force due to the solute transport is in opposite direction. As in the case 1, thermal effects drive the clockwise convection. Due to the solutal effects, weak counter-clockwise convection loops appear near the horizontal walls. These results are not in agreement with observational data [1,7] which show clockwise loops as in case 1.

The solute boundary layer thickness for case 2 is much thicker than case 1 (bottom panel). This may be caused by the increased mixing between the thermal and solute boundary layers at the vertical walls (bottom panel). A strong temperature inversion occurs in the upper right and lower left corner of the ampoule as in case 1.

Comparison with Measurement Data

Figure 3 shows the horizontal velocity (u), nondimensionalized temperature and solute distributions at the

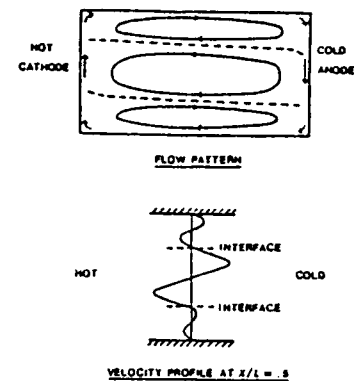
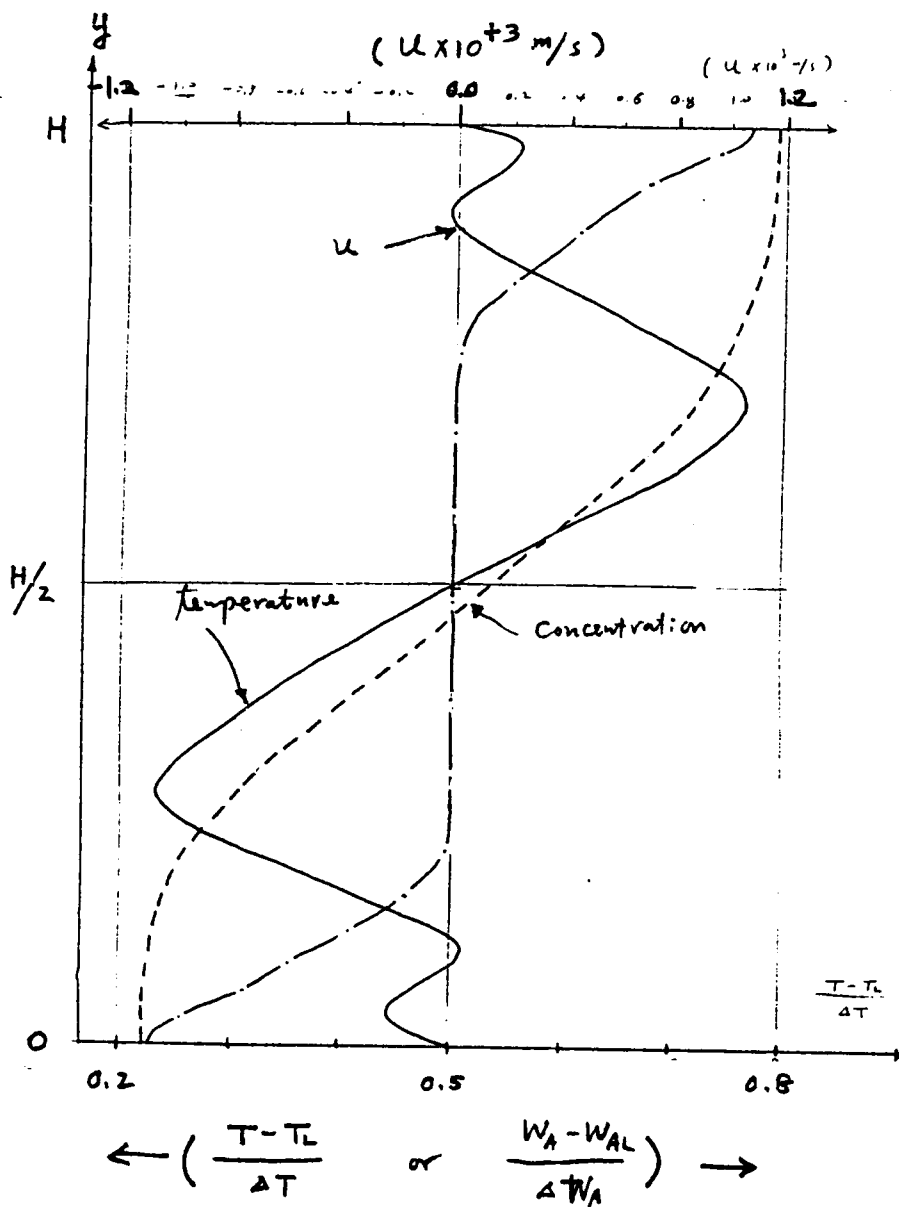


FIG. 2. Layered flow structure for cooperating case.

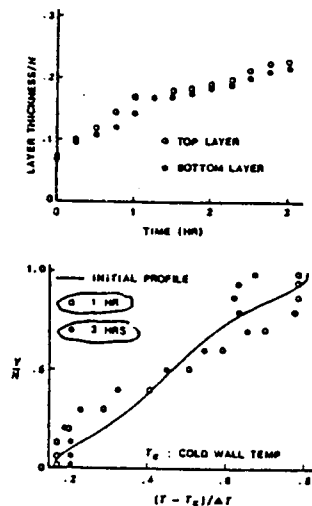


FIG. 4. Changes of layer thicknesses and temperature profile at mid-section for cooperating case with $Ar = 0.55$, $Gr_s = 1.0 \times 10^7$, $Gr_f = 5.4 \times 10^5$ and $N = 18.5$.

Figure 3. Vertical distribution of velocity, temperature and concentration at the center of the cavity for case 1 at $t=4000\text{sec}$. Experimental results taken from refs.[1,7] are shown for comparison.

ORIGINAL PAGE IS
OF POOR QUALITY

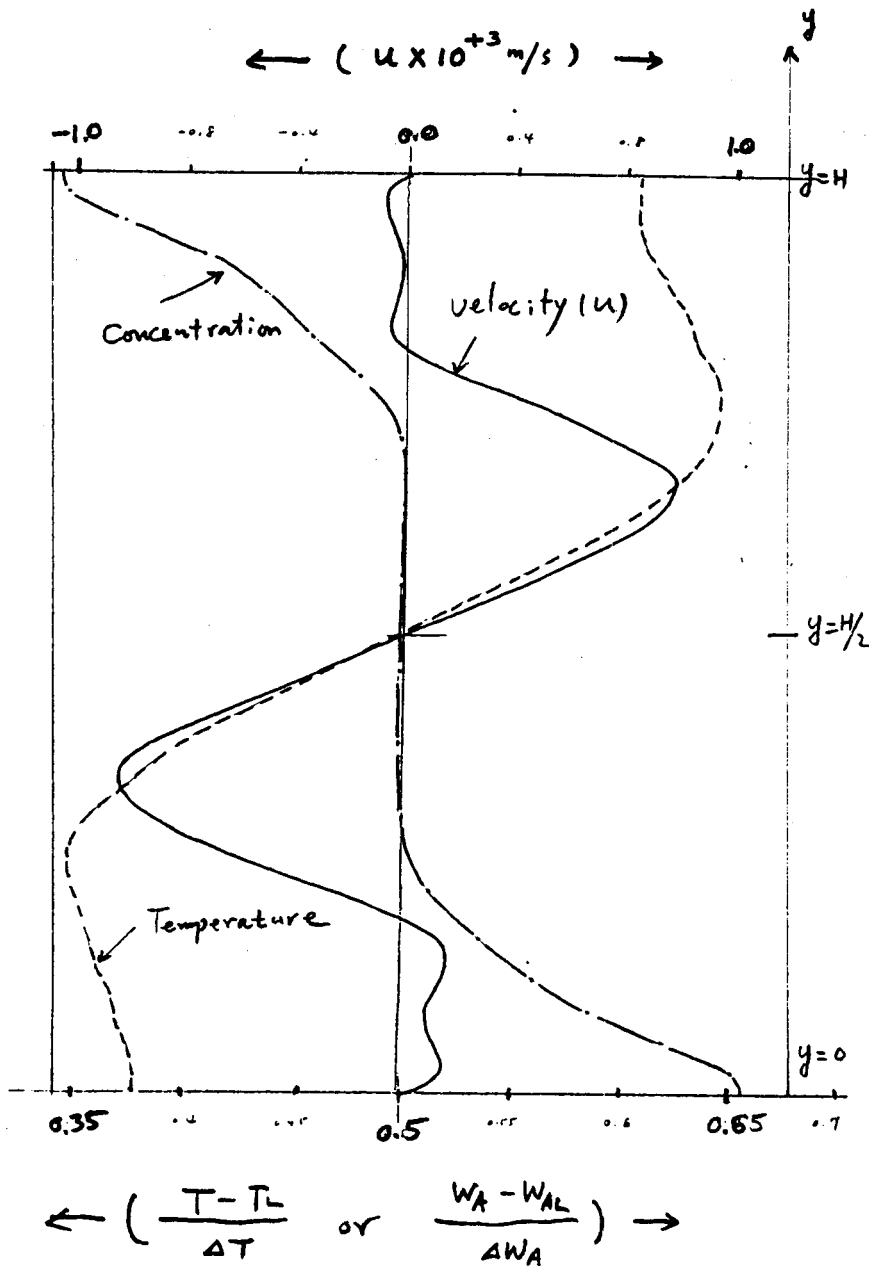


Figure 4. Vertical distribution of velocity, temperature and concentration at $x=L/2$ for case 2 at $t=4000$ sec. Experimental results for a similar case in refs.[1,7] are shown for comparison.

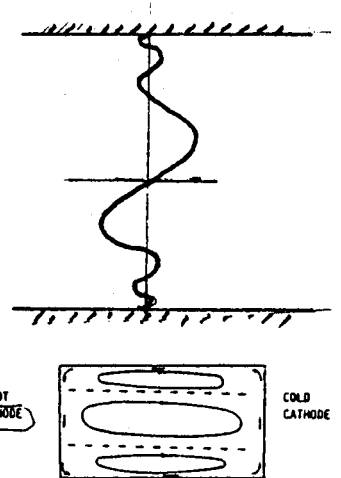


Fig. 16 Sketch of typical flow structure produced at a certain time, showing the layers formed by heating from side for an opposing case

$$\begin{aligned} Gr_1 &= 6.2 \times 10^5, Gr_2 = 5.8 \times 10^5 \\ A &= 0.55, Sc = 2.1 \times 10^3 \end{aligned}$$

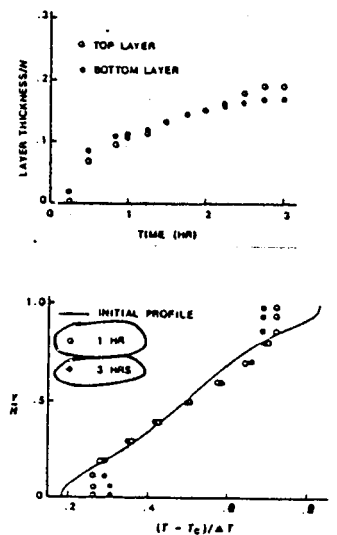


Fig. 7. Changes of layer thicknesses and temperature profile at mid-section for opposing case with $Ar = 0.55, Gr_1 = 1.0 \times 10^5, Gr_2 = 6.2 \times 10^5$ and $N = 16.2$.

center of the cavity as a function of vertical distance for case 1 at $t=4000$ sec. Experimental data from ref. [7] are also displayed for comparison. Good qualitative agreements are observed for the velocity and temperature fields. A direct comparison of concentration profiles is not possible since no measurement data is available. The thickness of the solute layer is about 20 % of the cavity height in the present results while it is about 15 % in the measurement.

Figure 4 shows the similar results for case 2 at $t=4000$ sec. Again experimental data with slightly different conditions [1,7] are shown for comparison. Experimental results show three convective loops rotating clockwise in addition to a thin counter-clockwise circulation due to opposing solutal convection. Numerical results however failed to show two clockwise convection loops near the horizontal surfaces. There exist, however, a crude resemblance between the velocity profiles. Temperature profiles obtained by the numerical simulation show extreme inversion while the measurement data shows almost uniform temperature in the horizontal solute boundary layer. Solute layer thickness from the present results is about 25 % cavity height while measurement reports only 12 %. Even though some of this discrepancy can be attributed to the difference in the buoyancy ratio used in the present study (18.6) and the measurement (16.2), no satisfactory explanation can be offered.

Average Heat and Mass Transfer

Average heat and mass transfer rate across the cavity are of engineering interest. They are commonly expressed in term of Nusselt number (Nu) and Sherwood number (Sh) defined by

$$Nu = \frac{\text{actual energy transfer by conduction and convection}}{\text{energy transfer by conduction alone}} \quad (9)$$

and

$$Sh = \frac{\text{actual mass transfer by diffusion and convection}}{\text{mass transfer by diffusion alone}}$$

For a transient heat and mass transfer, Nu and Sh evaluated at the different locations would give different values. They will be constant when a steady state is reached. Three locations, at the left wall ($x=0$), at the middle of the cavity ($x=L/2$) and at the right wall ($x=L$), are selected to evaluate Nu and Sh in the present study.

Table 3 summarizes the average Nu and Sh numbers evaluated at $x=0$ at the end of computations. Computations were terminated when the average Sh number variations at $x=0$ are relatively mild. Even though a true steady-state can not be reached for a very long time because of small mass diffusivity, measurement results [1,7] showed that the mass transfer rate reached a quasi-steady value within a relatively short time. For example, measurement

showed that limiting current (constant mass flux) occurs within 1800 sec for all cases they studied. Numerical results also showed such trend in average Sh numbers. The average Sh numbers in Table 3 represent the quasi-steady values.

Numerical results agree well with the measured Sh for all cases of aiding convection. Discrepancy between two results are within the uncertainty of measurements (10 %). As mentioned in a previous section, horizontal layering of solute near the top and bottom surfaces reduce the cavity height and consequently, thermal convection strength decreases. Therefore, the average Nu number continues to drop as time progresses unlike the average Sh number which reaches a quasi-steady value. Therefore, the average Nu numbers tabulated in Table 3 are not quasi-steady values. Similar trend in reduction of heat transfer by solute layering was also noted in ref.[14].

Table 3. Summary of average heat (Nu) and mass transfer (Sh) rate evaluated at $x=0$

Case	present		measurement (ref.7)	
	Nu	Sh	Nu	Sh
1	14.95	265.30	-	281.83
2	7.09	215.64	-	-
3	10.32	138.36	-	125.83
4	3.76	119.16	-	-
5	4.14	72.38	-	68.0

From the numerical results, it can be observed that the opposing buoyancy force has much more effects on the heat transfer rate than on the mass transfer rate. In an opposing convection, thermal convection is greatly reduced by the opposing solute buoyancy force and the heat transfer rate becomes smaller. Mass transfer occurs through a thin solute boundary layer at the vertical wall which changes less dramatically than the thermal boundary layer.

Figure 5 shows the momentum, thermal and solute boundary layer thickness at $x=0$, $y=H/2$. Momentum boundary layer is thicker than the thermal and the thermal boundary is much thicker than the solute boundary layer as expected since $Pr=7.07$ and $Sc=2140$.

By reducing the gravity, thermal and solute Gashof numbers are both reduced by the same amount and the average heat and mass transfer rates are equally affected as seen in Table 3. Reducing

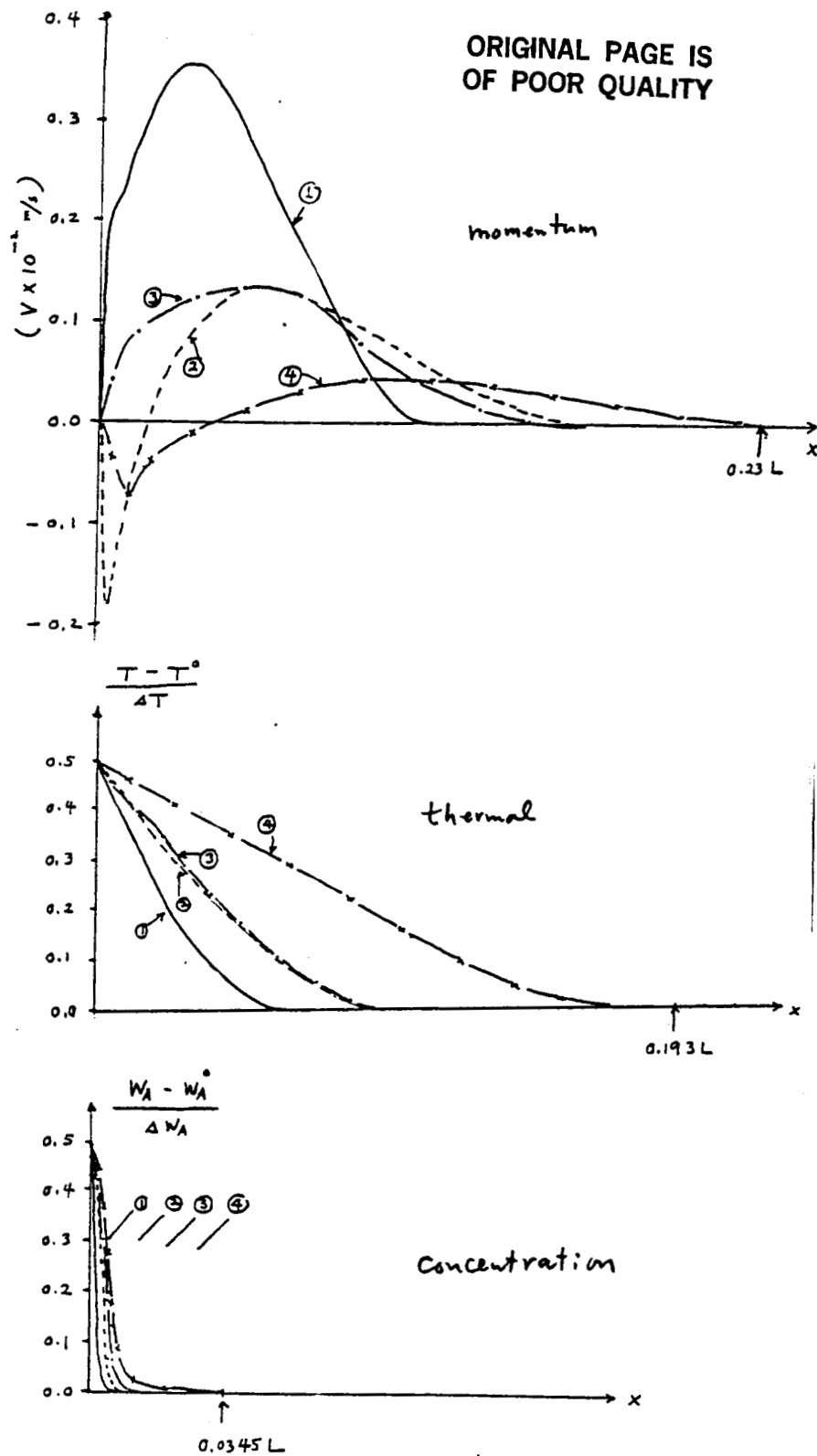


Figure 5. Momentum, thermal and concentration boundary layer thicknesses at the vertical wall at $y=H/2$ for case 1-4.

the gravity to a level of microgravity environment, say $10e-4g$, may not however eliminate the convection effects on the mass transfer. For example, reducing gravity to $0.1g$ reduces Sh by 48 % (from 265.3 to 138.36) and with $0.01g$, 46 % (from 138.36 to 74.0). This shows that the mass transfer is decreased by 50 % by decreasing the gravity ten times. At $1.0e-4g$, the average Sh may become 18.75. In order to have $Sh=1.0$, the gravity may have to be reduced to $1.0e-8g$.

Flow Instabilities

Figure 6 shows the transient Nu and Sh number variations for case 2 evaluated at $x=0$ and $x=L/2$. Results at $x=L$ are similar to those at $x=0$ and are not shown. Top panels show the mass transfer rate and the bottom panel shows the heat transfer rate.

Mass transfer evaluated at the wall drops rapidly at the beginning because the mass transfer occurs by diffusion and the solute gradient decreases. The mixture adjacent to the wall begins to move because of solute buoyancy force and the solute boundary layer develops. The concentration at the edge of solute boundary layer is the initial concentration in the ampoule. The mass transfer rate changes very little until convection initiated from the right wall reaches at about $t=370$ sec. At that time the concentration at the edge of the solute boundary layer begins to drop below the initial concentration and the average mass transfer rate starts to increase.

The mass transfer evaluated at the center remains zero until significant convection of solute occurs at the middle of the ampoule. Note that the mass transfer at the center is carried by convection and diffusion is almost negligible. Mass transfer at the center reaches the maximum when the horizontal velocities near the horizontal walls were the maximum. This occurred at about $t=370$ sec for case 2. Convection strength decreases continually but solute layer near the horizontal walls increases. This is the reason that the average mass transfer rate reaches a quasi-steady value (limiting current value in measurement) rather rapidly. In the present case, this value is reached at about 800 sec.

Mass transfer rate at the wall is seen to increase continuously at $t>800$ sec. This implies that the solute concentration at the edge of solute boundary layer is being lowered continuously. Mass transfer at the center remains at the quasi-steady value until about 1300 sec at which time oscillation begins. These oscillations persisted until the end of computation while mass transfer rate at the wall remains almost constant. There are two types of oscillations: one with a short oscillation period (about 15 sec) and the other with a relatively long period (about 150 sec).

The behavior of transient heat transfer rate at the wall

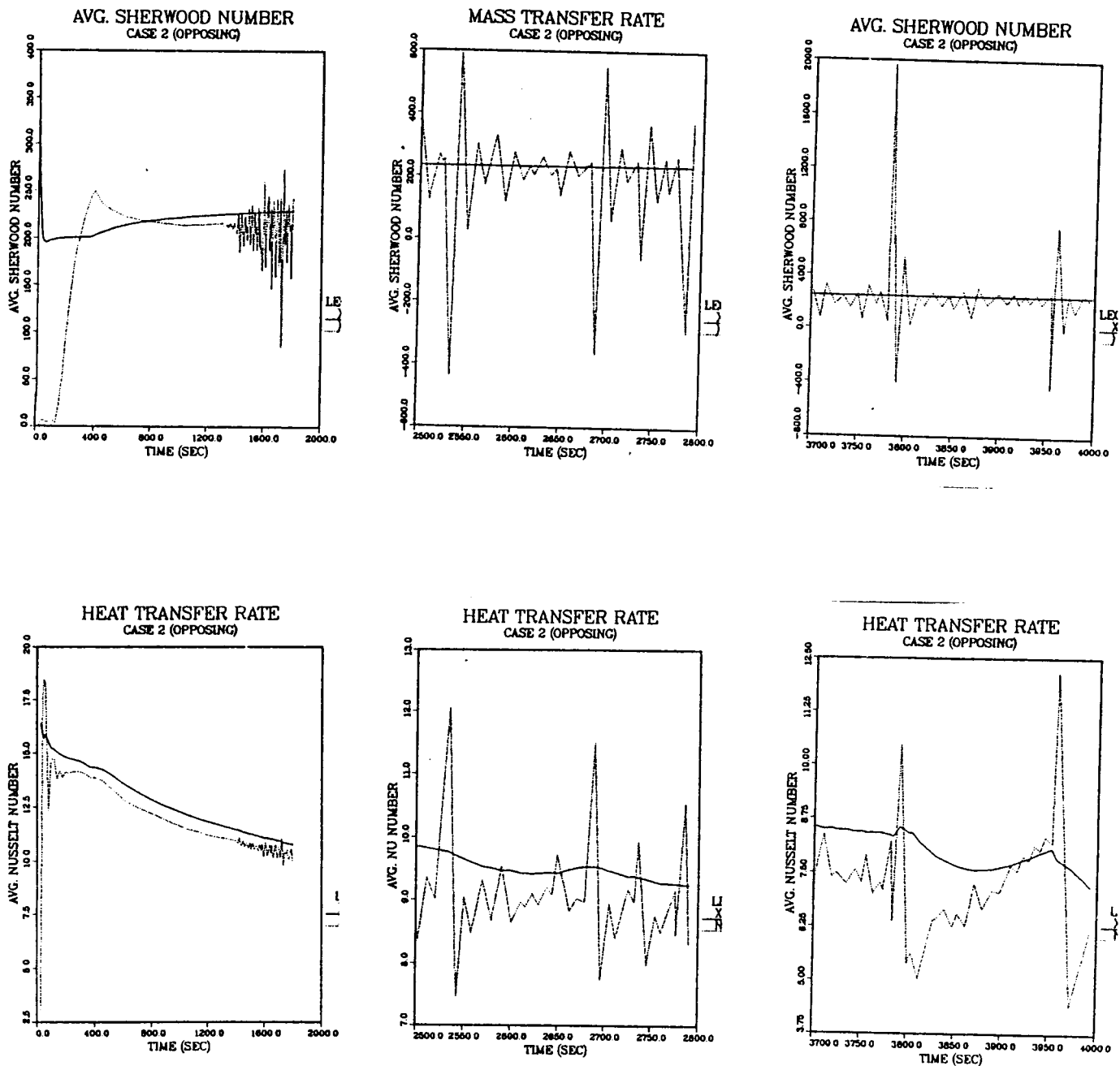


Figure 6. Transient average mass transfer (top panels) and average heat transfer (bottom panels) across the cavity for case 2 evaluated at $x=0$ (solid lines) and $x=L/2$ (broken lines).

and at the center can be explained in a similar manner. Since the thermal diffusivity is much larger than the mass diffusivity, thermal boundary layer development is much quicker and the thermal boundary layer thickness much larger (see Fig.5). As explained, the heat transfer rate does not reach a quasi-steady value. Instead, they decrease continuously as the effective height of the cavity decreases. Oscillation of Nu follows closely of those of Sh fluctuations.

Some of these oscillations are probably caused by the Brunt-Vaisala instability [15]. Fluid with a destabilizing temperature field (cold fluid on top of warm fluid) is unstable. By using the physical parameters found in the present case, the period of oscillation is found to be

$$\tilde{p} \approx 2\pi / \left\{ \frac{(\alpha \nu Gr_T Pr)^{1/2}}{\sqrt{1+A^2}} \left(\frac{1}{H^2} \right) \right\} \approx 10.77 \text{ sec} \quad (10)$$

Other type of short oscillation is caused by the interaction between a destabilizing thermal field and a stabilizing solute field. Using the formula given in ref. [13], the period of oscillation for a over-stable marginal stability for the present case is

$$\tilde{p} \approx 2\pi / \left[\frac{\nu^2}{3Sc^2 H^4} \left\{ \frac{Sc(Sc-Pr)}{Pr(Pr+1)} (Gr_s \cdot Pr) - \frac{27\pi^4}{4} \right\} \right]^{1/2} \approx 12.8 \text{ sec} \quad (11)$$

These two instability mechanisms are probably connected to the short period oscillation observed for Nu and Sh evaluated at the center.

The long period oscillation with large amplitude may be associated with a boundary layer instability at the vertical walls. Interaction among the momentum, energy and thermal buoyancy force are shown to produce such boundary layer instability [16,17]. Since the momentum boundary layer instability produces fluctuations in the velocity boundary layer thickness, thermal boundary thickness will also be affected for moderate Pr number flows. This is indeed the case. Nusselt number at the wall shows such fluctuations. No changes are observed for the Sh at the wall since the solute boundary layer thickness is too small ($Sc=2140$). Modifying the formula given in ref.[17] for the present case, the period of oscillation is estimated to be

$$\tilde{p} \approx 4\pi^{-1} \nu^{1/3} (g\alpha \Delta T_f)^{-2/3} Gr_s^{2/3} \approx 150 \text{ sec}, \quad (12)$$

where ΔT_r is the temperature difference across the thermal boundary layer and ζ_{r_f} is based on the momentum boundary layer thickness.

Similar oscillations of both short and long periods were observed in case 1 as well. Oscillation began at about $t=2500$ sec in case 1. The critical Grashof number without solutal effects is about 60000 [17]. Grashof number for case 1 and 2 are about 20000. Solute effects, therefore, seem to lower the critical Grashof number over which the velocity boundary layer becomes unstable.

For reduced gravity cases, no boundary layer instabilities were observed. This implies that the critical Grashof number is larger than 2000. The other short periods oscillations were present in reduced gravity cases. But the amplitude of oscillations were very small.

Interfacial Velocity Effects

It is assumed in the present analysis that the velocity at the vertical walls is zero. In reality, however, horizontal velocity due to diffusion of species A is not zero at these interface between the fluid and the solid. By taking the mass balance at the interface, it can be shown that [18] the horizontal velocity at $x=0$ and $x=L$ are given by

$$u(x=0, y) = - \frac{D_{AB}}{1 - w_A(0, y)} \frac{\partial w_A}{\partial x} \bigg|_{x=0, y} \quad (13)$$

and

$$u(x=L, y) = - \frac{D_{AB}}{1 - w_A(L, y)} \frac{\partial w_A}{\partial x} \bigg|_{x=L, y} \quad (14)$$

respectively.

To show the effects of interfacial velocity, the above velocity conditions were used for case 5 replacing the appropriate boundary conditions. The results showed that the average Sh at $x=0$ increased by only 2 points. This small change is expected since the interfacial velocity in a dilute solution is very small compared with the characteristic velocity in a moderately large Grashof number convection.

IV. CONCLUSIONS AND RECOMMENDATIONS

In conclusions, a two-dimensional time-dependent numerical analysis are used to investigate the thermal and solute interactions encountered in a solution growth of single crystals. Numerical results were compared with experimental data. Good quantitative agreements were obtained in term of average mass transfer rate. Good qualitative agreements of over-all convection fields were also observed.

It was observed that the thermal and solute fields become highly oscillatory when the Grashof number are large. Oscillations are probably caused by a number of instability mechanisms. Some of these instabilities were seen to disappear at the reduced gravity environments.

Transient temperature and solute distributions are found to be fairly non-uniform at the crystal growing surface ($x=L$). And the effects of convection on the solute transport may not be negligible even in the micro-gravity environment with $1.0e-4g$.

The following is recommended for the extension of the present study:

1. It is further needed to verify the capability of the numerical method used in the present study by simulating other experimental data. (Lack of satisfactory agreement between the numerical and experimental results for case 2 should be adequately explained.) One possible source of disagreement is the numerical diffusion. A higher order numerical approximation may be needed.
2. Analyse other solution growth situations with realistic physical and flow parameters, such as mixture properties, ampoule geometries and boundary conditions (including segregation effects).
3. Perform a fundamental numerical study on the instability mechanisms encountered over simple geometries such as vertical and horizontal planes.
4. Extend to three-dimensional cases if success is warranted.

REFERENCES

1. Ostrach, Simon, "Fluid Mechanics in Crystal Growth , 1982 Freeman Scholar Lecture," J. Fluid Eng., Vol. 105, 5-20, 1982.
2. Rosenberger, F., Fundamentals of Crystal Growth, Vol. 1, Macroscopic Equilibrium and Transport Concepts, p. 360, Springer, Berlin, 1979.
3. Pimpulkar, S. M., and Ostrach, S., "Convective Effects in Crystal Grown from Melt," J. Crystal Growth, 55, p.614, 1981.
4. Chang, C. J., and Brown, R. A., "Radial Segregation Induced by Natural Convection and Melt/Solid Interface Shape in Vertical Bridgman Growth," J. Crystal Growth, 63, p.343, 1983.
5. Murthy, J. Y., "A Numerical Simulation of Flow, Heat and Mass Transfer in A Floating Zone At High Rotational Reynolds Numbers," J. Crystal Growth, 83, p.23, 1987.
6. Extremet, G., Roux, B., Bontoux, P., and Elie, F., "Two-dimensional Model for Thermal and Solutal Convection in Multizone Physical Vapor Transport," J. Crystal Growth, 82, p.761, 1987.
7. Kamotani, Y., Wang, L. W., Ostrach, S., and Jiang H. D., "Experimental Study of Natural Convection in Shallow Enclosure with Horizontal Temperature and Concentration Gradients," Int. J. Heat, Mass Transfer, 28, p.165, 1985.
8. Patankar, S. V., Numerical Heat Transfer and Fluid Flow, McGraw-Hill, 1980.
9. Han, S. M., "A Generalized Implicit Finite Difference Method for Transient Analysis of Compressible and Incompressible Fluid Flow," in Numerical Methods for Fluid Transient Analysis, ASME FED-Vol.4, Martin and Chaudry(editors), p.17, 1983.
10. Van Doormal, J. P., and Raithby, G. D., "Enhancement of the SIMPLE Method for Predicting Incompressible Fluid Flows," Numerical Heat Transfer, Vol.7, p.147, 1984.
11. Han, S. S., and Schafer, C. F., "Transient Natural Convection Heat and Mass Transfer in A Rectangular Enclosure: A Numerical Analysis," in Development of Theoretical and Applied Mechanics, Vol. 14, Hackett and Wang (editors), 1988.
12. Turner, J. S., "Double-Diffusive Phenomena," Ann. Rev. Fluid Mech., Vol. 6, p.37, 1974.
13. Shirtcliffe, T. G. L., "An Experimental Investigation of Thermosolutal Convection at Marginal Stability," J. Fluid Mech., Vol.35, p.677, 1969.

14. Writz, R. A., "The Effect of Solute Layering on Lateral Heat Transfer in an Enclosure," Int. J. Heat, Mass Transfer, Vol. 20, p. 841, 1977.
15. Patterson, J., and Imberger, J., "Unsteady Natural Convection in a Rectangular Cavity," J. Fluid Mech., Vol. 100, p. 65, 1980.
16. Gebhart, B. "Instability, Transition, and Turbulence in Buoyancy-Induced Flows," Ann. Rev. Fluid Mech., Vol. 5, p. 213, 1973.
17. Carruthers, J. R., "Origin of Convective Temperature Oscillations in Crystal Growth Melts," J. Crystal Growth, 32, p. 13, 1976.
18. Greenwell, D. W., Markham, B. L., and Rosenberger, F., "Numerical Modeling of Diffusive Physical Vapor Transport in Cylindrical Ampoules," J. Crystal Growth, 51, p. 413, 1981.

N89-21742

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

OPTIMIZATION OF LARGE MATRIX CALCULATIONS FOR EXECUTION
ON THE CRAY X-MP VECTOR SUPERCOMPUTER

Prepared by:	Dr. William A. Hornfeck
Academic Rank:	Professor
University and Department:	Mississippi State University Electrical Engineering Department

NASA/MSFC	
Office:	Information Systems
Division:	Systems Development and Implementation
Branch:	Engineering Systems
MSFC Colleague:	Mr. Bobby C. Hodges
Date:	July 29, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

OPTIMIZATION
OF
LARGE MATRIX CALCULATIONS
FOR EXECUTION
ON THE CRAY X-MP
VECTOR SUPERCOMPUTER

by

William A. Hornfeck
Professor of Electrical Engineering
Mississippi State University

ABSTRACT

A considerable volume of large computational computer codes have been developed for NASA over the past twenty-five years. This code represents algorithms developed for machines of an earlier generation. With the emergence of the vector supercomputer as a viable, commercially available machine, an opportunity exists to evaluate optimization strategies to improve the efficiency of existing software. This result is primarily due to architectural differences in the latest generation of "large-scale" machines and the earlier, mostly uniprocessor, machines. This report describes a software package being used by NASA to perform computations on large matrices, and describes a strategy for conversion to the Cray X-MP vector supercomputer.

ACKNOWLEDGEMENTS

There were quite a large number of persons at MSFC who proved to be very capable and quite eager to provide assistance and motivation during the ten-week course of this study. Certainly my colleague in this effort, Mr. Bobby C. Hodges, has my sincere thanks. Thanks also to his fellow NASA Information System Office personnel John C. Lynn, Director, Sheila Fogle, Shirley Thompson, Carla Krivutza, and Joe Pollock. From NASA's Structures and Dynamics Laboratory, I want to thank John Admire and Dave McGhee who loaned their expertise and computer codes.

I am grateful to a cross-section of NASA Contractor personnel, all of whom were extremely cooperative: Dale Robertson and Deborah Hagar of Grumman Data Systems, Larry Hoelzeman of Cray Research, Karin Offik of New Technology Inc., and Les Wade of Boeing Computer Support Services.

Ernestine Cothran of the MSFC Director's Executive Staff and Dr. Mike Freeman of the University of Alabama performed an outstanding service in coordinating and administering all aspects of the Summer Faculty Program.

INTRODUCTION

The FORMA (Fortran Matrix Analysis) software package was developed by Martin-Marietta approximately twenty years ago. This package has been adapted by NASA for use by the Flight Dynamics Laboratory at MSFC in solving large structures response equations:

$$-W^2 M + S \phi = 0$$

Where ϕ = Mode

M = Mass Matrix

S = Stiffness Matrix

W = System Eigenvalues

$$L(T) = A \frac{d^2 X}{dT^2} + B \frac{dX}{dT} + CX + DF(T) + E$$

Where L = Load Matrix

A, ..., E = Constant

T = Time

X = Position

F = Forcing Function

and Maximum Dimensions = (12,000 X 12,000)

Typical Matrix Dimensions = (500 X 500)

Atypical Matrix Dimensions = (5,000 X 5,000)

Original FORMA codes were adapted for execution on the MSFC UNIVAC 1108 Multiprocessor. These codes have been "ported" to a next-generation UNIVAC machine, then the IBM 3084, and now the Cray X-MP. Conversions were accomplished in a minimum of time, but without attention to optimization strategies regarding the host machines. The Cray is particularly sensitive to vector constructs within programs.

OBJECTIVES

Develop and adapt specialized mathematical/engineering techniques or methodologies to the solution of scientific/engineering problems utilizing supercomputer technology. Mathematical analyses and modeling of large computerized programs will be performed and recommendations for optimizing the solutions will be formulated. Oral and written reports will be presented/developed on research activities and results.

THE COMPUTING ENVIRONMENT

The Engineering Analysis and Data System (EADS) provides the Cray user at MSFC with a front-end to the supercomputer mainframe. Jobs submitted to the Cray are submitted through EADS. Figure 1 shows the system configuration for EADS.

The portion of EADS which is important to Cray/FORMA users is shown in Figure 2. Also included as part of this figure are the three general areas of concern in optimization studies for codes executing on the Cray.

SEADS

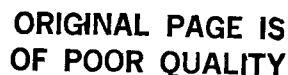
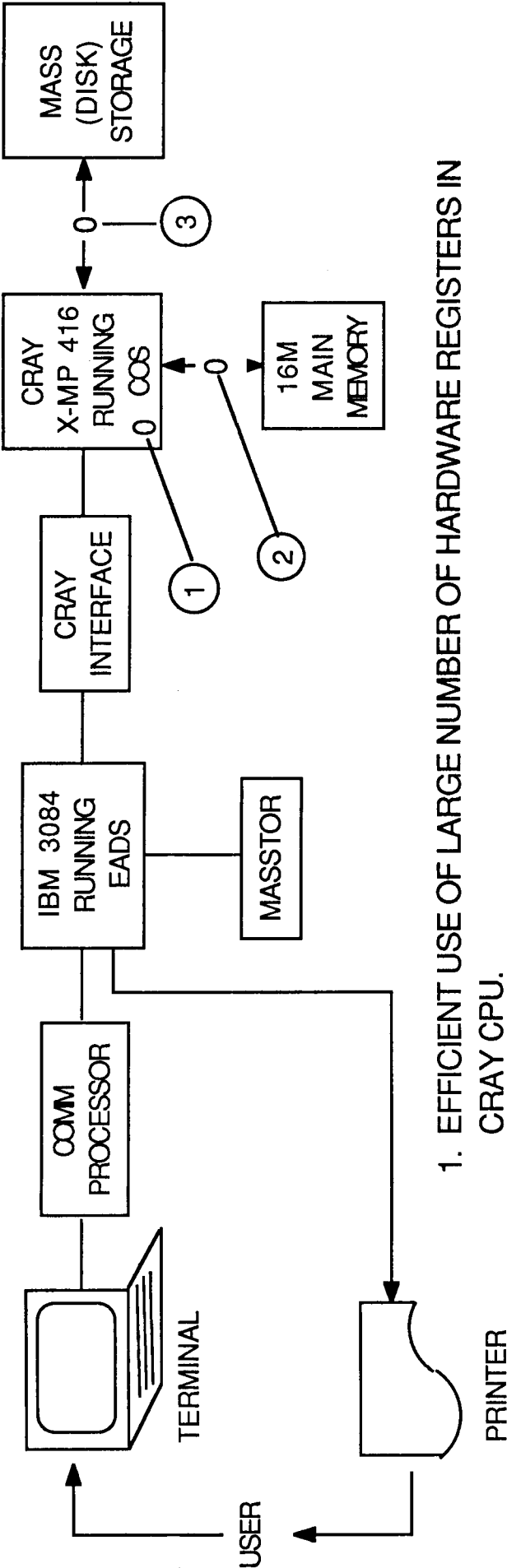


FIGURE 1. EADS SYSTEM

THE FORMA SYSTEM ENVIRONMENT



1. EFFICIENT USE OF LARGE NUMBER OF HARDWARE REGISTERS IN CRAY CPU.
2. REDUCE THE NEED FOR MAIN MEMORY TRANSFER DURING PROGRAM EXECUTION.
3. STREAMLINE THE DATA TRANSFERS WHICH MUST TAKE PLACE BETWEEN CRAY MAIN MEMORY AND DISK STORAGE.

FIGURE 2. FORMA ENVIRONMENT

OPTIMIZATION STUDIES

The FORMA (Fortran Matrix Analysis) software package consists of the following:

105 MATRIX ANALYSIS SUBPROGRAMS:

- o 42 Arithmetic Subprograms
- o 45 Matrix Manipulation Subprograms
- o 12 I/O Utility Subprograms
- o 6 System Utility Subprograms

The FORMA subroutines are characterized by the attributes listed here:

o MODULAR FORTRAN STRUCTURE

The average arithmetic routine is 180 statements

The average matrix manipulation routine is 80 statements

The average I/O utility routine is 30 statements

The average system utility routine is 10 statements

o ARITHMETIC STRUCTURE

Matrices as large as 12,000 X 12,000 are processed by using submatrices of dimension 60 X 60, plus residues

o SUBPROGRAM DEPENDENCIES

The average subprogram requires 5 arguments in call statement. The average subprogram call 3 other subprograms.

- o VECTORIZATION

All vectorization is presently the result of compiler-generated codes. The average subprogram contains approximately 2 vector loops set up in this fashion.

The optimization for vector processing will be very sensitive to the existing FORMA subprograms; however, the Cray X-MP architecture is equally important. Figure 3 shows the basic register configuration for the Cray X-MP. The references at the conclusion of this report provide detailed specifications on the architecture and COS operating system.

Of particular importance in the optimization process is the organization of the 8 64-word vector registers and associated vector functional units. The peak computing speeds achievable by the Cray are principally attributable to sustained vector computations.

The existing FORMA subprograms should be analyzed for the following optimization factors:

- o Subroutine/function calls
- o Loop indices and addressing of arrays
- o Order dependencies and recursions
- o use of scalars in do loops
- o Decision processed
- o Restructuring do loops
- o General rules

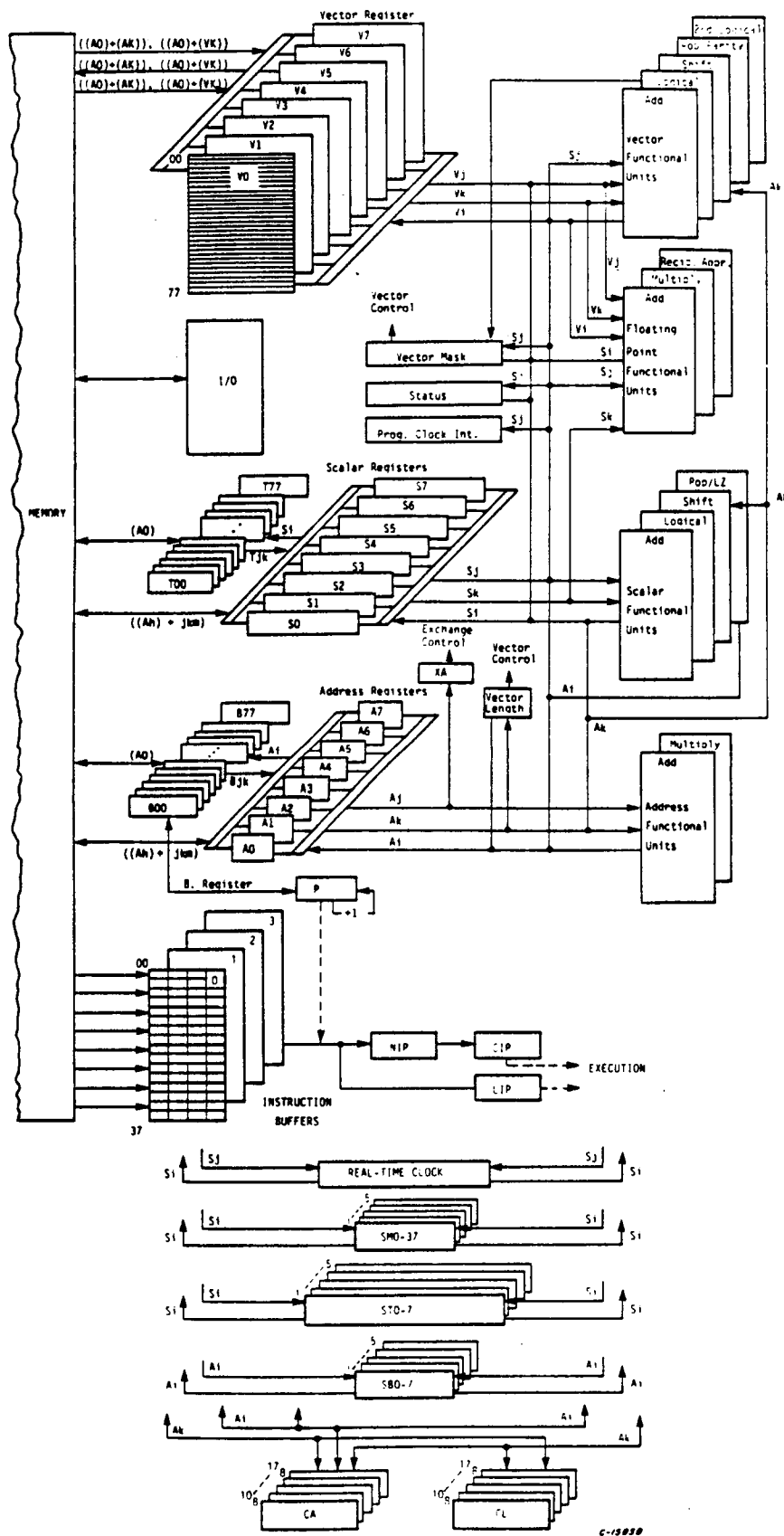


FIGURE 3. CRAY X-MP/4 BLOCK DIAGRAM

ORIGINAL PAGE IS
OF POOR QUALITY

Each of the optimization factors is now broken down into a more detailed list of do's and don'ts relative to vectorization:

CHECK GENERAL RULES

- o Avoid double precision;
- o use memory interleaving;
- o Avoid integer divides;
- o Use parentheses;
- o Avoid mixed mode expressions.

CHECK SUBROUTINE/FUNCTION CALLS

- o Isolate non-vectorizable function CALLS;
- o Separate D) loops for non-vector functions;
- o Remove (nonrecursive) SUBR CALLs from DO loops;
- o Use statement functions;
- o Convert function CALLs to user vector functions.

CHECK ORDER DEPENDENCIES-RECURSIONS:

- o Simple subscripts help compiler to recognize vectorizable loops;
- o Vectorize code on non-recursive loop indices;
- o Recognize order-dependencies--these are recursions which can be reordered to remove the dependence on order;
- o Truly recursive operations should be placed in separate DO loops;
- o Optimize when vectorize is not possible.

CHECK DECISION PROCESSES:

- o Remove loop-independent IF statements from DO loop;
- o Remove IF tests on loop indices and adjust loop bounds accordingly;
- o Create separate loops for "low-probability" decision statements involving loop indices;
- o Use temporary variable outside DO loop range for "low-probability: decision statements;
- o Avoid the computed GOTO;
- o IF-THEN-ELSE is not vectorizable;
- o Restructure conditional statements according to "density of the decision process";
- o Perform both halves of condition and then select proper results (mask undesirable ones);.

CHECK RESTRUCTURING DO LOOPS:

- o Even if additional calculations required, remove scalar statements from DO loops;
- o Use vector length of 64 whenever possible;
- o Make longer loops the innermost loops;
- o If possible, convert nested DO loops into a single DO loop;
- o Always combine DO loops of equal length;
- o "Unroll: small outer loops;
- o "Expand" small inner loops.

CHECK THE USE OF SCALARS IN DO LOOPS:

- o Check reduction functions, which result in scalars;
- o use MIN, MAX, IMIN, IMAX functions;
- o Check dot products, which result in scalars;
- o Use the SDOT functions;
- o Check matrix multiplication, which results in a reduction from 2 matrices to a single matrix;
- o Use matrix multiplication kernel which allows maximum vectorization (see example);
- o Convert scalar recursions to vector arrays;
- o Do not use loop indices in loop calculations.

CHECK LOOP INDICES AND ADDRESSING OF ARRAYS:

- o Check indirect addressing;
- o Avoid use of indirect addressing in generating more compact codes;
- o Use GATHER/SCATTER functions;
- o Sparse matrices are exception;
- o Whenever possible, repeated indices should have constant "stride";
- o No complicated expressions for loop indices;
- o Repeated memory references which differ by 8 or 16 locations can cause memory bank conflicts.

OPTIMIZATION STRATEGIES

There are several approaches to accomplishing the conversion of existing, non-vectorized computer codes to obtain more efficient Cray X-MP programs. In this section, a short-term strategy will be suggested and an example analysis will be discussed. In addition, a long-term conversion strategy will be outlined, along with a general optimization procedure.

Figure 4 is a flowchart of a short-term optimization procedure which addresses the conversion of more critical subprograms on a priority basis. This flowchart is specific to the FORMA software package, and when the procedure is followed for a typical job stream, we obtain the following results:

1. FORMA routines have been classified one time (this step not part of a loop) and documented, noting several key parameters and briefly describing function.
2. Typical job stream obtained from System Response Branch (ED22). This program calculates a response matrix and requires approximately 25 CPU-SEC to execute.
3. Flow trace utility provides the following statistics:

<u>Subprogram Name</u>	<u>% Run-Time</u>	<u>Subprogram Function</u>
RESPONS	2.33	Main Program
NTRANI	7.73	I/O Utility
NTRANR	11.00	I/O Utility
ZRDISK	3.78	I/O Utility
ZWDISK	1.80	I/O Utility
ZMULX1	35.63	$[Z] = [A] * [B] + [Z]$
ZMULT	28.60	$[Z] = [A] * [B]$
ZMAXMN	2.52	$r_{\max} = \max [R]$
SOLVEQ	1.86	$A \frac{d^2 X}{dt^2} + B \frac{dX}{dt} + CX = 0$
OTHER	4.75	36 other subprograms

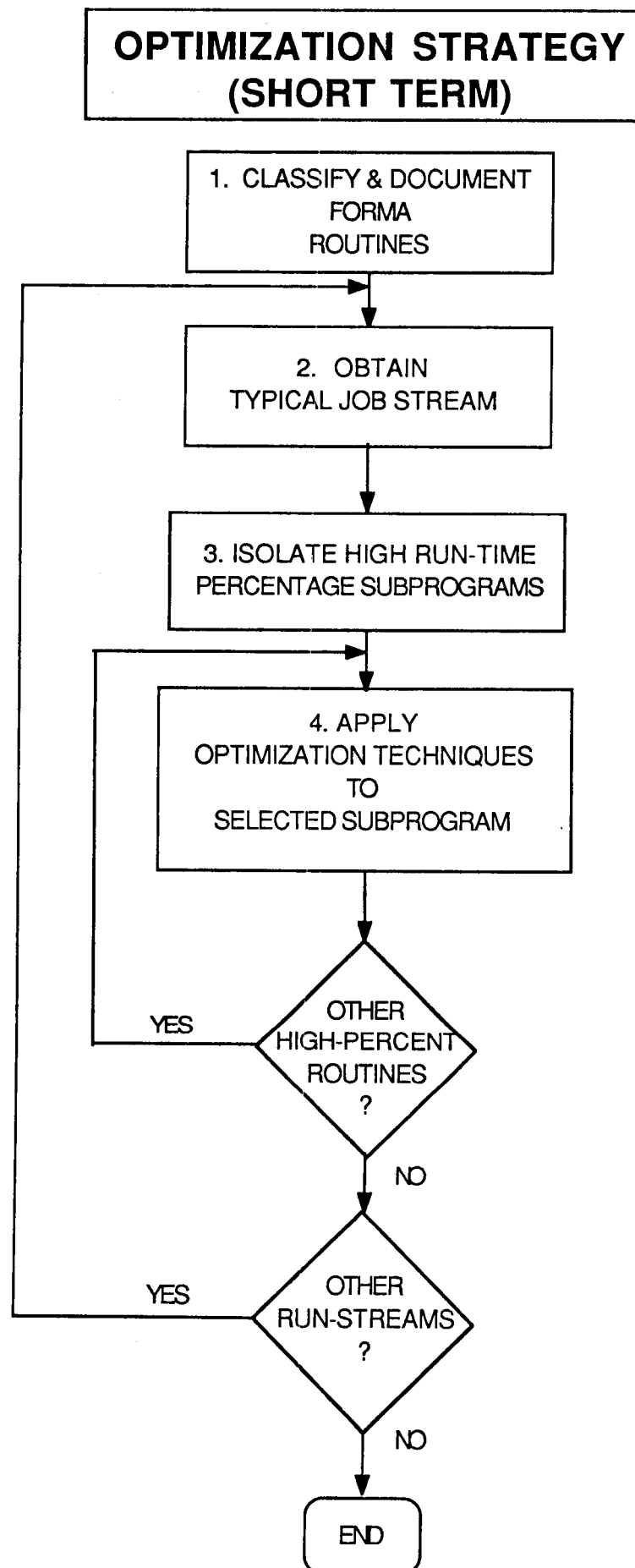


FIGURE 4. SHORT-TERM OPTIMIZATION STRATEGY

4. ZMULX1, ZMULT Optimization:

Since these are similar routines, optimization methods will be similar;

Restructure vector loops: one in each;

Isolate subroutine calls, especially I/O;

Use of scalars in DO loops;

Vectorize decision processes, if appropriate;

General rules.

5. We shall treat the discussion of block number 5 in the optimization strategy by showing a typical analysis process involving matrix multiplication. First, consider the "normal" matrix multiplication program segment:

```
DO 10 I= 1,N
DO 10 J= 1,N
A (I,J) = 0.0
DO 10 K= 1,N
10 A (I,J) = A(I,J) + B(I,K) * C (K,J)
```

Then consider a "better multiply kernel which allows the Cray compiler to set up more efficient vector calculations:

```
DO 9 J=1,N
DO 9 I=1,N
A (I,J)=0.0
9 CONTINUE
DO 10 K=1,N
DO 10 J=1,N
DO 10 I=1,N
A(I,J)=A(I,J) + B(I,K) * C(K,J)
10 CONTINUE
```

Notice that the vectorized code is not as compact, but it allows the Cray to perform two vector calculations at the innermost loop of both nested-DO's.

Figures 5 and 6 show the ZMULT and ZMULX1 routines which were found to be the highest-run-time subprograms in our typical run stream. The reader should compare the DO loop structure discussed above with these figures.

```

1. SUBROUTINE ZMULT1(NMSA,NMSB,NMSZ)
2. DOUBLE PRECISION SA,SB,SZ
3. DATA KRCPRT/60/
4. COMMON /LZ1/ INDA(204),INDRPA(200),MHA(10),SA(60,60)
5. COMMON /LZ2/ INDB(204),INDRPB(200),MHB(10),SB(60,60)
6. COMMON /LZ3/ INDZ(204),INDRPZ(200),MHZ(10),SZ(60,60)

C
C THIS IS A SPECIAL MULTIPLICATION SUBROUTINE DESIGNED TO BE USED
C BY ZTRAE. IT PERFORMS THE OPERATION:
C (A)*(B) + (Z) = (Z)
C THIS ROUTINE ALLOWS THE NUMBER OF ROWS IN (B) TO BE LESS
C THAN THE NUMBER OF COLUMNS IN (A).
C CALLS FORMA SUBROUTINES: CHKZER,DZERO,ZBEGIN,ZCLEAN,ZREDI,ZREDR,
C ZWRTI,ZWRTIR,ZZBOMB.
C DEVELOPED BY JOHN ADMIRE. MAY 1981.
C LAST REVISION BY JOHN ADMIRE. JAN 1984. (LEWIS CRAY)
C IMPLEMENTED ON IBM 3084 BY DAVID S. MCGHEE. MARCH 1986.

C SUBROUTINE ARGUMENTS (ALL INPUT)
C NMSA - PARTITION-LOGIC IDENT FOR MATRIX (A).
C NMSB - PARTITION-LOGIC IDENT FOR MATRIX (B).
C NMSZ - PARTITION-LOGIC IDENT FOR MATRIX (Z).
C
C NERROR EXPLANATION
C 1 = NUMBER OF COLUMNS IN (A) LESS THAN NUMBER OF ROWS IN (B).
C 2 = NUMBER OF ROWS IN (A) NOT EQUAL TO NUMBER OF ROWS IN (Z).
C 3 = NUMBER OF COLUMNS IN (B) NOT EQUAL TO NUMBER OF COLUMNS IN (Z).
C

7. CALL ZBEGIN(NMSA,NRA,NCA,NRPA,NCPA,NRLA,NCLA,INDA,MHA)
8. CALL ZBEGIN(NMSB,NRB,NCB,NRPB,NCPB,NRLB,NCLB,INDB,MHB)
9. CALL ZBEGIN(NMSZ,NRZ,NCZ,NRPZ,NCPZ,NRLZ,NCLZ,INDZ,MHZ)
10. NERROR=1
11. IF(NCA.LT. NRB) GO TO 999
12. NERROR=2
13. IF(NRA.NE. NRZ) GO TO 999
14. NERROR=3
15. IF(NCB.NE. NCZ) GO TO 999
16. DO 150 IRPA=1,NRPA
17. : NRSA=KRCPRT
18. : IF(IRPA.EQ. NRPA) NRSA=NRLA
19. : CALL ZREDI(INDRPA,200,INDA(IRPA))
20. : CALL ZREDI(INDRPZ,200,INDZ(IRPA))
21. : :-----
22. : : DO 140 JCPB=1,NCPB
23. : : NCSB=KRCPRT
24. : : IF(JCPB.EQ. NCPB) NCSB=NCLB
25. : : IF(INDRPZ(JCPB).LE. 0) GO TO 100
26. : : CALL ZREDR(SZ,KRCPRT*KRCPRT,INDRPZ(JCPB))
27. : : GO TO 110
28. : : 103 CALL DZERO(SZ,NRSA,NCSB,KRCPRT)
29. : : 110 CONTINUE
30. : : DO 130 IRPB=1,NRPB
31. : : NCSA=KRCPRT
32. : : IF(IRPB.EQ. NRPB) NCSA=NRLB
33. : : IF(INDRPA(IRPB).LE. 0) GO TO 130
34. : : CALL ZREDI(INDRPB,200,INDB(IRPB))
35. : : IF(INDRPB(JCPB).LE. 0) GO TO 130
36. : : CALL ZREDR(SA,KRCPRT*KRCPRT,INDRPA(IRPB))
37. : : CALL ZREDR(SB,KRCPRT*KRCPRT,INDRPB(JCPB))
38. : : :-----
39. : : DO 120 I=1,NRSA
40. : : : DO 120 J=1,NCSB
41. : : : : Rp- DO 120 L=1,NCSA
42. : : : : : Rp- 120 SZ(I,J)=SZ(I,J)+SA(I,L)*SB(L,J)
43. : : : : : 130 CONTINUE
44. : : : : CALL CHKZER(SZ,NRSA,NCSB,IFZERO,KRCPRT)
45. : : : : IF(IFZERO.LE. 0.AND. INDRPZ(JCPB).GT. 0)
46. : : : : *INDRPZ(JCPB)=-INDRPZ(JCPB)
47. : : : : IF(IFZERO.GT. 0)
48. : : : : *CALL ZWRTIR(SZ,KRCPRT*KRCPRT,INDRPZ(JCPB))
49. : : : : 140 CONTINUE
50. : : : : 150 CALL ZWRTI(INDRPZ,200,INDZ(IRPA))
51. : : : : CALL ZCLEAN(NMSZ,INDZ,MHZ)
52. : : : : RETURN
53. : : : : 999 CALL ZZBOMB(' ZMULT1 ',NERROR)
54. : : : : END
55. : : : : ONE LINE DO LOOP REPLACED AT SEQ. NO. 39, P= 2016

```

```

1.      SUBROUTINE ZMULT (NMSA,NMSB,NMSZ)
2.      DOUBLE PRECISION SA,SB,SZ,S,SS
3.      COMMON /LZ1/ INDA(204),INDRPA(200),MHA(10),SA(60,60)
4.      COMMON /LZ2/ INDB(204),INDRPB(200),MHB(10),SB(60,60)
5.      COMMON /LZ3/ INDZ(204),INDRPZ(200),MHZ(10),SZ(60,60)
6.      DIMENSION X(60,60)
7.      DATA KRCPRT/60/

      C
      C MATRIX MULTIPLICATION FOR PARTITION-LOGIC. (A) * (B) = (Z).
      C CALLS FORMA SUBROUTINES CHKZER,ZBEGIN,ZCLEAN,ZOPEN,ZREDI,
      C ZREDR,ZWRTI,ZWRTR,ZBOMB.
      C DEVELOPED BY RL WORLEN. AUGUST 1977.
      C LAST REVISION BY JOHN ADMIRE. FEB 1982.
      C IMPLEMENTED ON IBM 3084 BY DAVID S. MCGHEE. MARCH 1986.
      C
      C SUBROUTINE ARGUMENTS (ALL INPUT)
      C NMSA = PARTITION-LOGIC NAME FOR MATRIX (A).
      C NMSB = PARTITION-LOGIC NAME FOR MATRIX (B).
      C NMSZ = PARTITION-LOGIC NAME FOR MATRIX (Z).
      C
      C ERROR EXPLANATION
      C I = MATRICES (A) AND (B) ARE NOT COMPATIBLE SIZE.
      C
      C READ MATRIX (A) HEADER.
      C CALL ZBEGIN(NMSA,NRA,NCA,NRPA,NCPA,NRLA,NCLA,INDA,MHA)
      C
      C READ MATRIX (B) HEADER.
      C CALL ZBEGIN(NMSB,NRB,NCB,NRPB,NCPB,NRLB,NCLB,INDB,MHB)
      C
      C CHECK (A) AND (B) MATRICES FOR SIZE COMPATIBILITY.
      C
10.      NERROR=
11.      IF (NCA .NE. NRB) GO TO 999
      C
      C FORM MATRIX (Z) HEADER.
      C NRZ=NRA
      C NCZ=NCB
12.      CALL ZOPEN(NMSZ,NRZ,NCZ,NRPZ,NCPZ,NRLZ,NCLZ,INDZ,MHZ)
13.
14.      C
      C MULTIPLY MATRICES (A) AND (B).
15.      DO 29 IRPA=1,NRPA
16.      :      NRSA=KRCPRT
17.      :      IF(IRPA.EQ. NRPA) NRSA=NRLA
18.      :      CALL ZREDI(INDRPA,200,INDA(IRPA))
19.      :      CALL ZREDI(INDRPZ,200,INDZ(IRPA))
20.      :      DO 25 JCPB=1,NCPB
21.      :      :      NCSB=KRCPRT
22.      :      :      IF(JCPB.EQ. NCPB) NCSB=NCLB
23.      :      :      CALL ZREDR(SZ,KRCPRT*KRCPRT,INDRPZ(JCPB))
24.      :      :      IFSZ=0
25.      :      :      DO 24 JCPA=1,NCPA
26.      :      :      :      NCSA=KRCPRT
27.      :      :      :      IF(JCPA.EQ. NCPA) NCSA=NCLA
28.      :      :      :      IF(INDRPA(JCPA) .LE. 0) GO TO 24
29.      :      :      :      CALL ZREDI(INDRPA,200,INDB(JCPA))
30.      :      :      :      IF(INDRPB(JCPB) .LE. 0) GO TO 24
31.      :      :      :      IFSZ=1
32.      :      :      :      CALL ZREDR(SA,KRCPRT*KRCPRT,INDRPA(JCPA))
33.      :      :      :      CALL ZREDR(SB,KRCPRT*KRCPRT,INDRPB(JCPB))
      C
34.      :      :      :      IF(NRSA.NE.KRCPRT .OR. NCSA.NE.KRCPRT .OR. NCSB.NE.KRCPRT) THEN
35.      :      :      :      :      DO 5000 J=1,NCSB
36.      :      :      :      :      :      DO 5000 K=1,NCSA
37.      :      :      :      :      :      :      V--
38.      :      :      :      :      :      :      DO 5000 I=1,NRSA
39.      :      :      :      :      :      :      :      SZ(I,J)=SZ(I,J)+SA(I,K)*SB(K,J)
40.      :      :      :      :      :      :      :      V
41.      :      :      :      :      :      :      :      5000 CONTINUE
42.      :      :      :      :      :      :      :      ELSE
43.      :      :      :      :      :      :      :      CALL MXM(SA,KRCPRT,SB,KRCPRT,X,KRCPRT)
44.      :      :      :      :      :      :      :      DO 5010 J=1,NCSB
45.      :      :      :      :      :      :      :      :      V--
46.      :      :      :      :      :      :      :      DO 5010 I=1,NRSA
47.      :      :      :      :      :      :      :      :      :      SZ(I,J)=SZ(I,J)+X(I,J)
48.      :      :      :      :      :      :      :      :      V
49.      :      :      :      :      :      :      :      :      5010 CONTINUE
50.      :      :      :      :      :      :      :      :      END IF
51.      :      :      :      :      :      :      :      C
52.      :      :      :      :      :      :      :      24 CONTINUE
53.      :      :      :      :      :      :      :      IF(IFSZ.EQ. 0) GO TO 26
54.      :      :      :      :      :      :      :      CALL CHKZER(SZ,NRSA,NCSB,IFZERO,KRCPRT)
55.      :      :      :      :      :      :      :      IF(IFZERO.EQ. 0) GO TO 26
56.      :      :      :      :      :      :      :      CALL ZWRTR(SZ,KRCPRT*KRCPRT,INDRPZ(JCPB))
57.      :      :      :      :      :      :      :      26 CONTINUE
      C
      C 29 CALL ZWRTI(INDRPZ,200,INDZ(IRPA))
      C CALL ZCLEAN(NMSZ,INDZ,MHZ)
      C
      C RETURN
      C

```

FIGURE 6. ZMULT SOURCE CODE

ORIGINAL PAGE IS
OF POOR QUALITY

The reader should also note that the Cray compiler has provided printout information showing all program loops, which are very important in the vectorization process. The compiler also marks each loop to inform the user of the vectorization which can be obtained, i.e., fully vectorized, conditionally vectorized, short vector loop, or a vector loop replaced by a subroutine call.

In examining Figures 5 and 6, it should be noted that, even for highly modular programs, the application of all vectorization rules which have been pointed out is a very tedious process. The vectorizing compiler provided by Cray, CFT or CFT77, performs well in finding vector constructs; however, it cannot perform as well as the vector programmer who carefully examines and optimizes codes to fully exploit the X-MP architecture. The following estimates conclude this example by calculating overall run-time improvement for RESPON if the stated levels of improvement are achieved for subprograms:

Estimate 25% improvement in ZMULX1

Estimate 25% improvement in ZMULT

Estimate 15% improvement in the other six predominant subroutines

This yields and estimated overall improvement of

$$(0.25) (0.64) + (0.15) (0.29) \approx 0.20,$$

or 20% improvement in a typical run stream.

Figure 7 shows a long-term strategy which could be employed if a complete conversion to vectorized code is justifiable for the FORMA package. This flow chart represents a procedure which would be a greater expense and requires more time, but which would yield a thorough redesign of the software.

A general optimization strategy is shown by the flow chart of Figure 8. This procedure is independent of the specific software package under consideration. Note that the procedure would require the implementation of general purpose test and data generation programs to thoroughly test vectorization strategies.

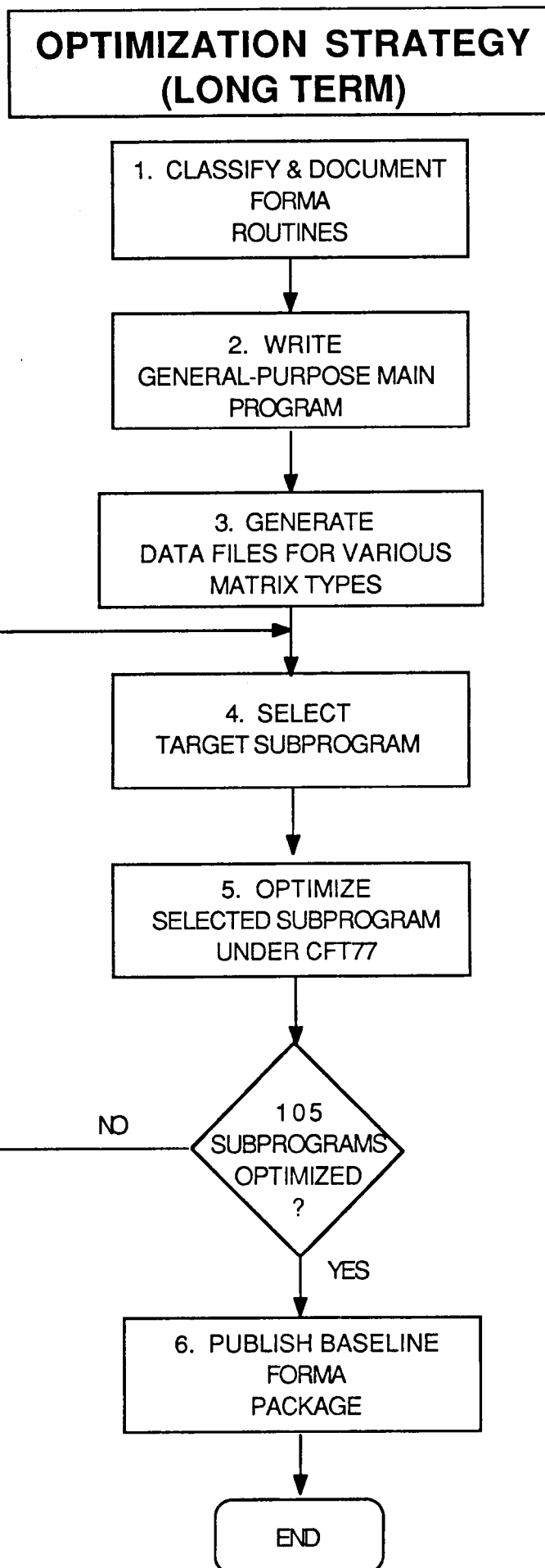


FIGURE 7. LONG-TERM STRATEGY

A GENERAL OPTIMIZATION STRATEGY

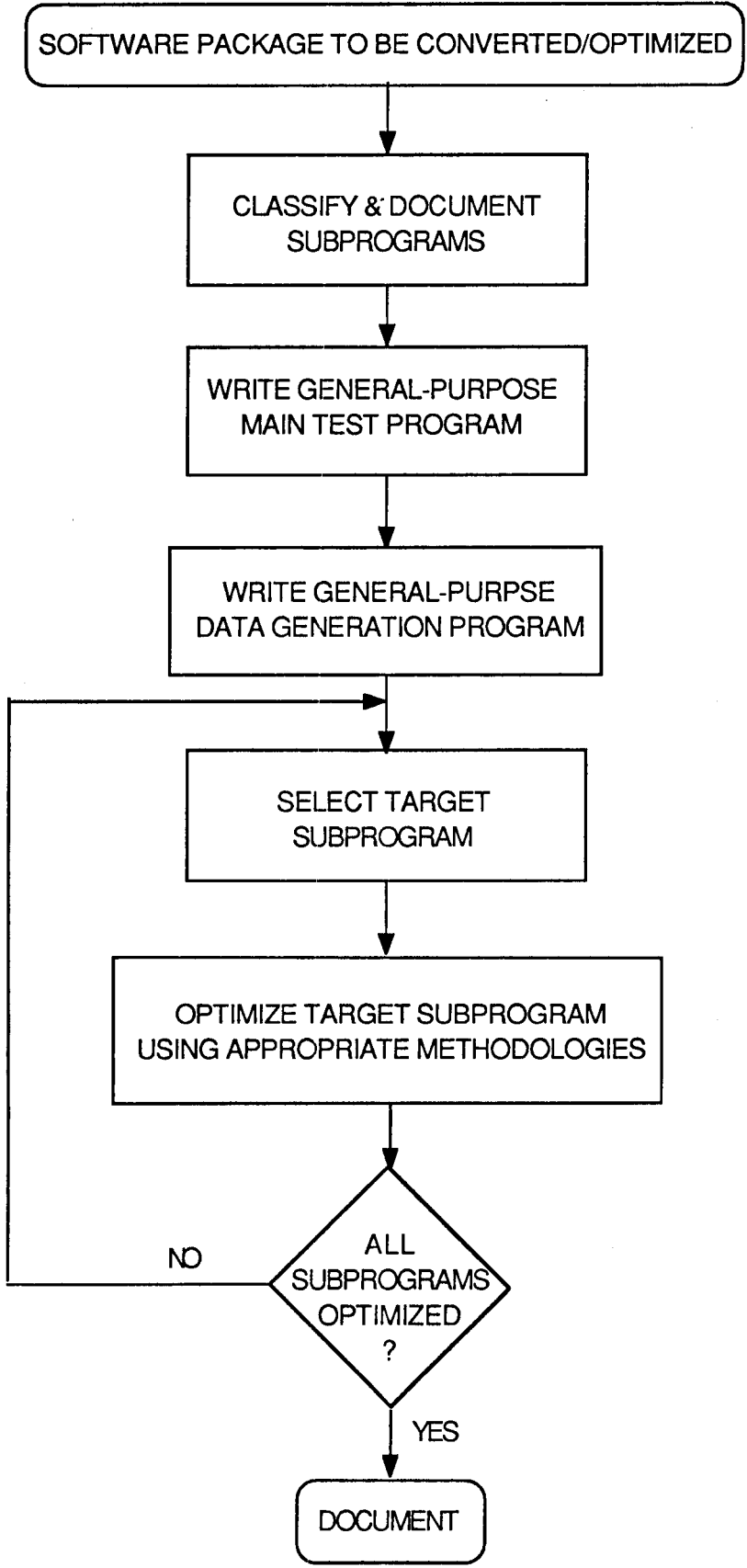


FIGURE 8. GENERAL STRATEGY
XVI-19

CONCLUSIONS AND FUTURE DIRECTION

Optimization of computer programs to achieve highly vectorized codes is a very exacting and time-consuming process. It is very much labor-intensive and it requires highly skilled personnel. On the other hand, these are rather costly attributes that must be balanced against the fact that software such as the FORMA routines are long-term investments. There are high initial costs associated with the optimization process, but there are long-term advantages to reducing CPU-minutes for frequently used programs.

The FORMA software package would be an excellent candidate for long-term optimization procedures. If this is done, several key areas would need to be addressed. These are:

- o The CFT77 compiler should be used in generating object code. In doing this, compiled codes should be compared with previous compilations to ensure the integrity of the compile process.
- o I/O utility routines are not particularly good candidates for optimization. However, these are frequently used routines and unique I/O speed-up features on the Cray should be investigated. These would include BUFFER IN/BUFFER OUT and unformatted I/O.
- o Custom performance monitoring routines should be implemented. These could provide users with a means to easily monitor performance enhancements and to monitor any difference in results obtained.
- o The optimization techniques which are effective tend to be reusable; that is, once learned or recognized, the same techniques can generally be applied a number of times in a given software package. Therefore, the more effective vectorization techniques should be well documented, including applicable performance statistics.

The Cray X-MP at NASA/MSFC represents a significant investment in high-performance computing technology. As such, resources to support this machine are critical. Those personnel writing new programs for the Cray X-MP should be well-versed in good vectorization techniques. In addition, permanent staff with in-depth knowledge of vectorization tools and techniques is important to the effective use of the present machine, as well as future upgrades and next-generation machines.

REFERENCES

FORTTRAN Programming on Cray Computers, Student Workbook,
Pacific-Sierra Corp.

Cray X-MP and Cray-1 Computer Systems, Programmer's Library
Reference Manual, Cray Research, Inc., 1987.

Cray X-MP and Cray-1 Computer Systems, FORTRAN (CFT)
Reference Manual, Cray Research, Inc., 1986.

EADS User's Guide, Grumman Data Systems, Doc. No. GDS-MSFC-
001, 1987.

Cray X-MP and Cray-1 Computer Systems COS Version 1 Reference
Manual and COS Version 1 Ready Reference Manual, Cray
Research, Inc., 1987.

Computer Architecture and Parallel Processing, Hwang and
Briggs, McGraw-Hill, 1986.

"Modal Analysis of Structures by an Iterative Rayleigh - Ritz
Technique", NASA TM X-64528, 1970.

N89-21743

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

INVESTIGATION OF PRACTICAL APPLICATIONS OF H_{∞} CONTROL THEORY
TO THE DESIGN OF CONTROL SYSTEMS FOR
LARGE SPACE STRUCTURES

Prepared by:	R. Dennis Irwin
Academic Rank:	Assistant Professor
University and Department:	Ohio University Department of Electrical and Computer Engineering
NASA/MSFC:	
Laboratory:	Structures and Dynamics
Division:	Control Systems
Branch:	Pointing Control Systems
MSFC Colleague:	Henry Waites
Date:	September 13, 1988
Contract:	NGT 01-002-099 The University of Alabama

INVESTIGATION OF PRACTICAL APPLICATIONS OF H_∞ CONTROL THEORY
TO THE DESIGN OF CONTROL SYSTEMS FOR
LARGE SPACE STRUCTURES

by

R. Dennis Irwin
Assistant Professor of Electrical and Computer Engineering
Ohio University
Athens, Ohio

ABSTRACT

The work documented here involves the investigation of the applicability of H_∞ control theory to the unique problems of large space structure (LSS) control. A complete evaluation of any technique as a candidate for large space structure control involves (1) analytical evaluation, (2) algorithmic evaluation, (3) evaluation via simulation studies, and (4) experimental evaluation. This report documents the results of analytical and algorithmic evaluations.

The analytical evaluation involves the determination of the appropriateness of the underlying assumptions inherent in the H_∞ theory, the determination of the capability of the H_∞ theory to achieve the design goals likely to be imposed on an LSS control design, and the identification of any LSS specific simplifications or complications of the theory. The results of the analytical evaluation are presented in the form of a tutorial on the subject of H_∞ control theory with the LSS control designer in mind.

The algorithmic evaluation of H_∞ for LSS control pertains to the identification of general, high level algorithms for effecting the application of H_∞ to LSS control problems, the identification of specific, numerically reliable algorithms necessary for a computer implementation of the general algorithms, the recommendation of a flexible software system for implementing the H_∞ design steps, and ultimately the actual development of the necessary computer code. The results and status of the algorithmic evaluation are presented.

Finally, the state of the art in H_∞ applications is summarized with a brief outline of the most promising areas of current research. Recommendations on further work are included, with emphasis on the LSS control problem.

INTRODUCTION

The original concept of an H_∞ performance criterion is that of Zames [1], who introduced the concept of minimum sensitivity and provided a solution in the SISO case. However it is highly probable that the need for such an optimization criterion has at various times been recognized by frequency domain oriented control engineers for quite some time. In any event, Zames motivation appears to be due to the recognition by the late 1970's that the by then well established LQG approach to control design suffered from fundamental limitations with regard to robustness(sensitivity). The fact that LQG based designs are inherently non-robust was quite surprising due to the fact that full state feedback quadratic optimal control designs have long been known to possess excellent robustness properties in the SISO case [2] and Safanov and Athans [3] had recently proved a corresponding result for MIMO full state feedback designs. The unexpected and unwanted result that the addition of a Kalman filter estimator can lead to arbitrarily small stability margins even in the SISO case was dramatically illustrated by an example of Doyle [4]. This led to abandonment of indirect approaches to achieving robustness and ultimately to the direct solution of the robustness optimization (minimum sensitivity, H_∞) problem by Zames.

The solution to the more difficult MIMO H_∞ optimization problem was obtained first by Francis, Helton and Zames [5] in 1984, who credit Cheng and Pearson [6] with an independent solution. Whatever the origin of the final solution, the stabilizing controller parametrization of Desour et al.[7] was the breakthrough that turned the H_∞ optimization problem into a practical control design approach, as opposed to a mathematical exercise in well established interpolation theory. In fact, it is expected that the mathematical tables will now be turned and, rather than the controller parametrization yielding practical applications of known mathematical theory, it will now motivate research into new mathematical optimization problems. This is the philosophy behind Vidyasagar's book on the factorization approach to control system synthesis [8].

While the "complete" solution to the original H_∞ optimization problems exists, work continues on the problem of controller order reduction, the complexity of the required algorithms, the possible equivalence to certain restricted classes of LQG problems(frequency weighted approaches, in particular), and the problem of simultaneously achieving both performance and robustness.

The approach taken here is pragmatic. H_∞ will be evaluated based on what can be done at the present time, with emphasis on the LSS problem, in regard to both the analytical issues of

applicability and the practical issues of software development. An outline of the H_∞ philosophy and theory is presented in a tutorial fashion. General algorithms that are well established are then presented and the actual numerical algorithms required are outlined in detail. Finally, possible design approaches appropriate for the LSS problem are outlined and design studies are suggested.

H_∞ THEORY

A Standard Problem

Although the original H_∞ problems were minimum sensitivity problems, a number of classical control design problems can be cast in the same form. The standard problem formalization is a way of casting any one of the problems of disturbance attenuation, command tracking, and robust stabilization into the form of a disturbance attenuation problem. The general form of the standard problem is given in block diagram form in Figure 1, where $W(s)$ is a vector of command and/or disturbance inputs, $Z(s)$ is a vector of regulated outputs which may be either tracking errors or actual regulated quantities, $U(s)$ is a vector of controller outputs, and $Y(s)$ is a vector of available measurements.

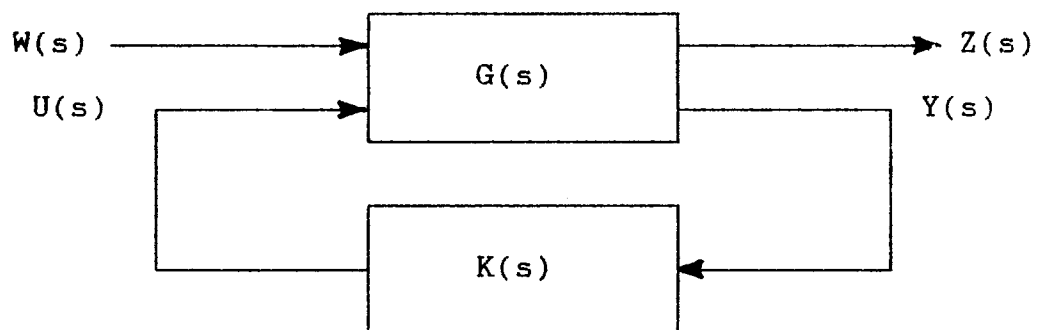


Figure 1

The relationships between the various signals are

$$Z = G_{11}W + G_{12}U, \quad (1)$$

$$Y = G_{21}W + G_{22}U, \quad (2)$$

and

$$U = KY. \quad (3)$$



Figure 2

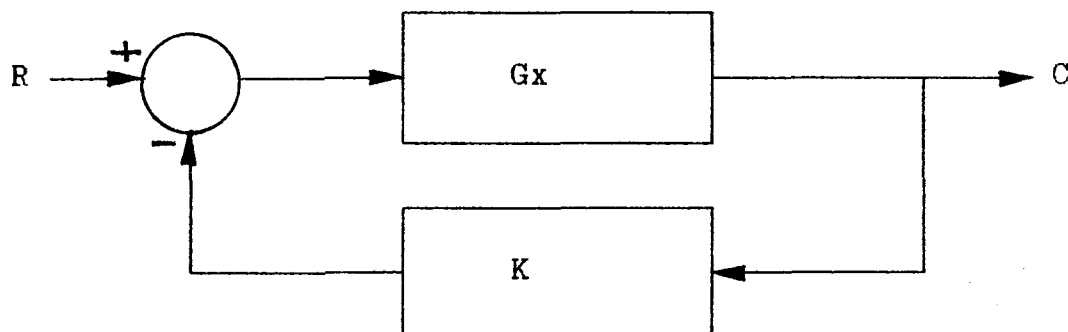


Figure 3

In Figure 3, the following definitions are made:

$$Z = R - C$$

$$W = R$$

$$U = KC$$

$$Y = C$$

Then

$$Z = W - Gx(W - U) = (1 - Gx)W + GxU$$

$$Y = Gx(W - U)$$

$$U = KY$$

which are in the form of a standard problem.

Mathematical Preliminaries

Before continuing with the parametrization of all stabilizing controllers, some quick definitions are needed.

DEFINITION

The space RH_∞ (scalar case) consists of all those rational functions with real coefficients which are stable and $\|G(\omega)\| < \infty$. For functions in RH_∞ , the norm can be calculated via

$$\|F\|_\infty = \sup\{|F(jw)| : w \text{ real}\}. \quad (4)$$

DEFINITION

The space RH_∞ (multivariable case) consists of all those matrices whose entries are stable rational functions with real coefficients and $\|G(\omega)\|_2 < \infty$. For matrices in RH_∞ the norm can be calculated as

$$\|G\|_\infty = \sup_w \{ \|G(jw)\|_2 \} \quad (5)$$

DEFINITION

The space RH_2 (scalar case) consists of all those rational functions with real coefficients which are stable and $|G(\omega)| = 0$. For functions in RH_2 , the norm can be calculated via

$$\|F\|_2 = \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} |F(jw)|^2 dw \right]^{1/2} \quad (6)$$

DEFINITION

The space RH_2 (multivariable case) consists of all those vectors whose entries are stable rational functions with real coefficients and whose elements are in RH_2 . For vectors in RH_2 the norm can be calculated as

$$\|F\|_2 = \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \|F(jw)\|_2^2 dw \right]^{1/2} \quad (7)$$

Simply stated, RH_∞ is the space of stable transfer functions (either scalar or multivariable) and RH_2 is the space of finite energy signals (scalar or vector valued). In H_∞ optimal control, the optimality criterion for choosing the controller K is to minimize

$$\| T \|_{\infty} = \sup_w \{ \| T(jw) \|_2 \}$$

where $T(s)$ is the closed loop transfer function from $W(s)$ to $Z(s)$ in the standard problem. The significance of this optimality criterion stems from the fact that

$$\| T \|_{\infty} = \sup \{ \| TW \|_2 \mid \| W \|_2 = 1 \} \quad (8)$$

which simply means that our goal is to minimize the output energy of the standard system for the "worst" input of unit energy. This goal is in contrast to the usual quadratic optimality criterion in which the goal is to minimize the output energy for a single, particular input.

The fact that H_{∞} optimization is not appropriate for all control design problems is seldom pointed out. However, in cases in which the system inputs (disturbances) are well known and the system model is also well known, there is little justification for the additional complexity of H_{∞} optimal control design. Indeed, in this case the usual quadratic optimal control approach should yield better performance than that of an H_{∞} approach.

Stabilizing Controller Parametrization

The mathematical theory on which H_{∞} optimization is based requires that the optimality criterion be stated in a form that requires only the choice between a set of transfer functions or matrices in RH_{∞} . However, once the optimal transfer function out of this set is chosen, it must in some way yield a controller which also internally stabilizes the system. The stabilizing controller parametrization accomplishes this goal by providing a one to one correspondence between elements of RH_{∞} and all stabilizing controllers.

The controller parametrization is based on factorization theory in RH_{∞} . A simplified derivation in the scalar case can be found in Irwin [9]; the complete derivation in the general case can be found in Vidyasagar [8] or Francis [10].

DEFINITION

A doubly co-prime factorization of a transfer function matrix G is a set of matrices in RH_{∞} which satisfy

$$G = NM^{-1} = M_1^{-1}N_1 \quad (9)$$

$$\begin{bmatrix} X_1 & -Y_1 \\ -N_1 & M_1 \end{bmatrix} \begin{bmatrix} M & Y \\ N & X \end{bmatrix} = I \quad (10)$$

It turns out that a doubly co-prime factorization of G_{22} is all that is needed to parametrize both the set of all stabilizing controllers and all stable transfer functions from W to Z . It is assumed that stabilizing G_{22} is equivalent to stabilizing G in the standard problem. The calculation of a doubly co-prime factorization is addressed in Nett et al. [11].

The end result of the controller parametrization is that all controllers which stabilize the system of the standard problem can be expressed as

$$\begin{aligned} K &= (Y - MQ)(X - NQ)^{-1} \\ &= (X_1 - QN_1)^{-1}(Y_1 - QM_1). \end{aligned} \quad (11)$$

where Q is some matrix in RH_∞ .

The parametrization of Equation 11 has significance well beyond the H_∞ optimal control problem. For example, it provides a simple stability check for a given candidate controller K ; simply solve Equation 11 for Q and if Q is in RH_∞ the controller stabilizes the system and if Q is not in RH_∞ the controller does not stabilize the system.

The closed loop transfer function matrix of the system of the standard problem can also be parametrized with Q as the parameter:

$$Z = [T_1 - T_2QT_3]W \quad (12)$$

where

$$T_1 = G_{11} + G_{12}MY_1G_{21} \quad (13)$$

$$T_2 = G_{12}M \quad (14)$$

$$T_3 = M_1G_{21}. \quad (15)$$

For systems which are open loop stable a valid doubly co-prime factorization is

$$Y = 0$$

$$Y_1 = 0$$

$$X = I$$

$$X_1 = I$$

$$M = I$$

$$M_1 = I$$

$$N = G_{22}$$

$$N_1 = G_{22}$$

so that a stabilizing controller is given by

$$K = -Q(I - G_{22}Q)^{-1}$$

and the closed loop transfer function matrix parametrization is

$$T_1 = G_{11}$$

$$T_2 = G_{12}$$

$$T_3 = G_{21}.$$

Notice that there is almost no work involved in parametrizing the problem when the system is open loop stable. In LSS problems such a situation can often be arranged via low bandwidth pre-compensation of the rigid body behavior. However, this fact does nothing to simplify the H_∞ optimization process.

The standard H_∞ optimization problem can be stated as finding a matrix Q in RH_∞ such that

$$\| T_1 - T_2QT_3 \|_\infty$$

is minimized. This is sometimes called the model matching problem. Once such a Q is found, Equation 11 can be used to find a controller which simultaneously achieves the optimization and stabilizes the system.

A few comments are in order regarding why the H_∞ problem is

difficult. The reasons are most easily seen by considering only those situations in which G is open loop stable and examining situations for which the equation

$$T_1 - T_2 Q T_3 = 0$$

can be solved. First, T_2 and T_3 must have left and right inverses, respectively, which implies that the system must have at least as many regulated quantities as controller outputs and at least as many external inputs as measurements. Neither of these conditions is overly restrictive as they are both equivalent to requiring that the system not have too many sensors or too many effectors. Then

$$Q = T_2^{-1} T_1 T_3^{-1}.$$

However, to be able to find a stabilizing controller, Q must be in RH_∞ , which implies that T_2^{-1} and T_3^{-1} must be in RH_∞ . This is equivalent to saying that T_2 and T_3 must have no transmission zeros in the right half of the complex plane and that they be non-strictly proper. In other words, T_2 and T_3 must be minimum phase and the resulting inverses must be realizable. The case in which they are non-minimum phase is the usual problem considered in H_∞ optimization theory. The case in which they are strictly proper can be dealt with via weighting matrices and therefore yields sub-optimal designs.

Solution of the Model Matching Problem

The theory of the H_∞ optimization step is quite advanced and involved in the multivariable case. Probably the best way to approach a presentation of the solution to derive it for the much simpler SISO case and then simply present the generalizations required for the solution in the MIMO case. A fairly simple SISO derivation using well known system theoretic concepts is contained in [9]. Another SISO derivation is contained in Francis [10], which also gives an authoritative proof (but not really a derivation) for the MIMO case. Various other approaches are documented in Cheng and Pearson [6], Safanov and Verma [12], Kwakernaak [13], and Grimble [14]. The approach here is to present the required steps in the solution in the form of an algorithm taken from Francis [10] with comments relating to the differences between the SISO and MIMO cases.

As mentioned previously, the major problem involved with

the model matching problem is that the inverses of T_2 and T_3 may not exist as matrices in RH_∞ . The consequence is that the optimality criterion is non-zero. The theory of H_∞ optimization tells how to find a Q in RH_∞ which achieves the minimum value. Although the theory is general enough to handle non-square, singular T_2 and T_3 , in the interest of clarity the following explanation and algorithm deals only with the square and nonsingular case. The completely general theory can be found in Francis [10] and a more compact explanation without proofs can be found in Irwin [9].

We begin with the problem of minimizing

$$T_1 - T_2 Q T_2 T_3$$

where T_1 , T_2 , and T_3 are in RH_∞ . This problem is converted to another equivalent problem by using inner-outer factorizations of T_2 and T_3 . An inner-outer factorization of a matrix G is two matrices G_i and G_o such that

$$G = G_i G_o \quad (16)$$

where

$$G_i^{-1} G_i = I \quad (17)$$

and G_o is right invertible with the right inverse in RH_∞ . In the case G is square, invertible, and is in RH_∞ , G_o is nonsingular in RH_∞ . A co-inner-outer factorization of G is given by

$$G = G_o G_{ci} \quad (18)$$

where

G_o^T is right invertible in RH_∞

and

$$(G_{ci}^T)^{-1} (G_{ci}^T) = I. \quad (19)$$

The first step in the H_∞ optimization solution is to find

an inner-outer factorization of T_2 and a co-inner-outer factorization of T_3 :

$$T_2 = T_{2i}T_{2o} \quad (20)$$

$$T_3 = T_{3co}T_{3ci}. \quad (21)$$

A description of a reliable state-space approach to calculating inner-outer factorizations can be found in Francis [10]. The closed loop transfer function matrix can then be written as

$$T_1 = T_{2i}T_{2o}QT_{3co}T_{3ci}. \quad (22)$$

Since inner and co-inner factors have unity norm by definition it is true that

$$\|T_1 - T_{2i}T_{2o}QT_{3co}T_{3ci}\|_{\infty} = \|T_{2i}^{-1}T_1T_{3ci}^{-1} - T_{2o}QT_{3co}\|_{\infty}. \quad (23)$$

Letting

$$R = T_{2i}^{-1}T_1T_{3ci}^{-1} \quad (24)$$

and

$$X = T_{2o}QT_{3co} \quad (25)$$

then

$$\|T_1 - T_2QT_3\|_{\infty} = \|R - X\|_{\infty}. \quad (26)$$

Now if X in RH_{∞} can be found to minimize the norm, then Q in RH_{∞} can be found by solving Equation 25, since T_{2o} and T_{3co} are invertible in RH_{∞} . Still more work is involved, however, before the application of the fundamental theory for finding such an X .

The next step is to find a minimal antistable/stable decomposition such that

$$R = R_1 + R_2 \quad (27)$$

where R_1 is completely unstable (R_1^{-} is in RH^∞) and strictly proper and R_2 is in H^∞ . The stable part of R is trivially included in X by letting

$$X = X_1 + X_2 \quad (28)$$

where

$$X_2 = R_2. \quad (29)$$

Then we have

$$\| T_1 - T_2 Q T_3 \|_\infty = \| R_1 - X_1 \|_\infty \quad (30)$$

where R_1 is antistable and X_1 is stable. This means that the H^∞ problem reduces to the problem of finding the closest stable matrix to a fixed antistable matrix. Finding a fixed antistable matrix R_1 such that Equation 30 holds is the main added difficulty incurred for the case of singular T_2 and T_3 . However the iterative procedure required obscures the central issues and is not included, since no additional algorithms or theories are required. The calculation of the antistable/stable decomposition of R is potentially the least reliable step in the calculation of an H^∞ optimal controller, since the corresponding state space representation of R must be converted to nearly block Jordan form; this process is infamous for its numerical instabilities.

The next step in the process of calculating Q is to find α , the minimum value of the optimality criterion:

$$\alpha = \min \| T_1 - T_2 Q T_3 \|_\infty = \min \| R_1 - X_1 \|_\infty. \quad (31)$$

The value of α can be calculated by

$$\alpha^2 = \max \{ \text{eigenvalues of } W_c W_o \} \quad (32)$$

where W_c and W_o are the controllability and observability grammians of any state-space realization of R_1 (A, B, C) and can be calculated by solving the Liapunov equations

$$AW_c + W_c A^T = BB^T \quad (33)$$

$$A^T W_o + W_o A = CTC. \quad (34)$$

From Equation 31 we also have

$$\min \| R_1/(\alpha+\beta) - X_1/(\alpha+\beta) \|_\infty = \alpha/(\alpha+\beta) \quad (35)$$

where $\beta > 0$.

Let

$$R' = R_1/(\alpha+\beta) \quad (36)$$

and

$$X' = X_1/(\alpha+\beta). \quad (37)$$

By Equation 35

$$\alpha/(\alpha+\beta) \leq \min \| R' - X' \|_\infty < 1. \quad (38)$$

Equation 38 seems rather unnecessary, but it turns out that it is easier to find X' such that

$$\| R' - X' \|_\infty \leq 1 \quad (39)$$

than it is to find X_1 in the original problem equal to a specified value. In any case when we find X' which satisfies Equation 39 we have

$$\alpha/(\alpha+\beta) \leq \min \| R' - X' \|_\infty \leq \| R' - X' \|_\infty \quad (40)$$

and since β is arbitrary we can find X' as close to the optimal as is desired.

The next, and final, step in the optimization procedure is to find X' such that with R' given in Equation 36, the relationship of Equation 39 holds. To begin, let (A, B, C) be a realization of R' and set

$$N = (I - W_o W_c)^{-1}. \quad (41)$$

$$L_1 = (A, N^T, C, 0) \quad (42)$$

$$L_2 = (-A^T, N W_o B, B^T, I). \quad (43)$$

Then an X' (not the only one) which satisfies Equation 39 is given by

$$X' = R' - L_1 L_2^{-1} \quad (44)$$

and the H_∞ optimization problem is solved, at least to an arbitrarily tight tolerance. The preceding development can be summarized as an algorithm:

Step 1

Cast the problem at hand in the form of the standard problem of Equations 1 - 3.

Step 2

Find a doubly co-prime factorization of G_{22} which satisfies Equations 9 and 10.

Step 3

Parameterize the closed loop transfer function matrix by using the results of Step 2 to find T_1 , T_2 , and T_3 of Equations 13 - 15.

Step 4

Find inner-outer and co-inner-outer factorizations of T_2 and T_3 .

Step 5

Use Equations 24 and 25 to find R and X .

Step 6

Find an antistable/stable decomposition of R , obtaining R_1 , R_2 , and X_2 .

Step 7

Calculate α , the optimal norm, using Equations 32 - 34.

Step 8

Choose a $\beta > 0$ to achieve the desired tolerance in Equation 40.

Step 9

Calculate R' from Equation 36.

Step 10

Calculate X' from Equations 41 - 44.

Step 11

Calculate X_1 from Equation 37 and X from Equation 28.

Step 12

Calculate Q by solving Equation 25.

Step 13

Calculate the H_∞ optimal controller from Equation 11.

Although the algorithm is stated in rather concise form, most of the separate steps involve a series of calculations themselves. The software requirements are discussed in the next section.

SOFTWARE SYSTEM AND ALGORITHM SUMMARY

Due to the complexity of the algorithms for accomplishing H_w control designs, a preliminary software system is under development. The purpose of the preliminary software is to lend flexibility and reliability to the software development process. The philosophy on which the software system is based is to develop a relatively small number of independent commands accessible via the particular operating system in use. Data is shared between these different low-level commands via a "state space system" representation which is sufficiently flexible to allow for all types of numerical data storage and which relies on the sequential file capabilities of the particular computer. An added benefit of this approach is that internal memory requirements are those of each particular module, rather than the memory required for the total design process. This approach is currently implemented on an MSDOS based 32-bit COMPAQ computer with Microsoft Fortran. However, the approach is fully consistent with any operating system and Fortran compiler capable of providing an interface between the operating system command line interpreter and Fortran program modules. This is certainly the case for MSDOS and UNIX based operating systems and can at the very least be modified for use on DEC VMS based machines. Of course, once design methods for H_w control become well established, the preliminary system can be converted to a more conventional single-module form, with the added advantage of the ability to directly evaluate memory requirements based on the experience of the preliminary system.

Data Representation

The only representation for shared data between modules is that of state space system representations stored in sequential files on system disk space. This representation allows for the storage of general linear continuous and discrete-time systems, two-dimensional array data, one-dimensional array data, and scalar data. For example, the system

$$x = Ax + Bu$$

$$y = Cx + Du$$

would be stored in a file named "system1" in the following form:

1st Record # inputs(m), # outputs(p), # states(n)
Next n Records rows of A
Next n Records rows of B
Next p Records rows of C
Next p Records rows of D.

An $i \times j$ matrix Z would be stored in file "Z" as:

1st Record j, i, 0
Next i Records rows of Z

so that a matrix is stored in the form of a system with j inputs, i outputs, and 0 states. A scalar quantity is stored as a 0 state system with 1 input and 1 output.

Low Level Modules

Low level modules are kept as simple as possible to allow for simplified debugging and to minimize internal storage. Typically, these modules do not destroy the contents of the input files. A list of the low level modules which exist at this time and their functions follows. Linear algebraic routines are taken, when possible, from LINPACK and EISPACK.

SADD A B C

Takes the systems contained in A and B, adds them as if their respective transfer function matrices were added, and places the result in the file C.

SMULT A B C

Takes the systems contained in A and B, multiplies them as if their respective transfer function matrices were multiplied, and places the result in the file C.

SINV A C

Takes the system A, inverts it as if its transfer function

matrix were inverted, and places the result in file C.
Limited to invertible systems (non-strictly proper).

SCOM A B C D S

Takes individual matrix elements of a system representation A, B, C, D in separate files and places them in a single system file format in S.

SSEP S A B C D

Takes the system representation S and places the matrices A, B, C, D in separate files.

SSTRAN S1 T S2

Applies the similarity transform T to S1 and places the result in S2.

SHTRANS S1 S2

Performs the state space operations on S1 which are equivalent to forming $G^T(-s)$ and places the result in S2.

ZCRIC A B Q R Z

Forms the "Hamiltonian" or "Z-matrix" required for Laub's Riccati equation solution [15]. A, B, Q, R are the equation coefficient matrices and Z is the resulting Hamiltonian matrix. (continuous form)

ZCLIA A Q Z

Same function as ZCRIC except for the solution of a Liapunov equation. (continuous form)

CSCHURS Z W1 W2

Forms an orthogonal similarity transformation which is used in Laub's method for solving Riccati and Liapunov equations. W1 and W2 are matrices formed from blocks of the transformation.

RSCHUR A T C

Transforms the matrix A to real schur form (block upper diagonal) via an orthogonal transformation T.

STRANS S1 S2

Takes the transpose of (the transfer function matrix of) S1 and places the result in S2

Low level modules still to be developed are:

ASDECOMP R R1 R2

Given a system R finds R1 antistable and R2 stable such that $R = R1 + R2$. (this requires what is almost a transformation to Jordan canonical form; an original algorithm has been developed to avoid many of the numerical instabilities, based on a method by Golub and Van Loan [16] for transforming a matrix to a special block diagonal form.)

MINIMAL S1 S2

Finds a minimal representation of the system S1 and places the result in file S2. (current plans are to use the "staircase" algorithm of Mayne [17]; if this algorithm proves unreliable a method due to Davison et. al [18] will be used. Davison's method is based on random choices of feedback to identify unobservable and/or uncontrollable modes.)

MSQRT A C

Finds a square root of a positive definite matrix A and places the result in file C

INFNORM S1 SCALAR

Finds the infinity norm of the system S1 and places the result in the file SCALAR. This involves first calculating the frequency response matrix of S1, and then calculating the maximum singular value at each frequency. This is the most numerically intensive step in the software system. An extremely efficient algorithm by Laub [20] will be used to calculate the frequency response matrix.

High Level Modules

High level modules will generally consist of batch-type files which contain a group of low level module commands. The full H_∞ design system will in turn be comprised of a combination of the low level and high level modules. Most of the high level modules are currently under development. A partial list of the required modules and their functions follows.

RICCATI AND LIAPUNOV EQUATION SOLUTIONS (existing)

Name: CRICSOL, CLIASOL

Other modules required:

ZCRIC, ZCLIA, CSCHURS, SMUL
SINV

DOUBLY CO-PRIME FACTORIZATIONS

Name: DCOFAC

Other modules required:

SADD, SMULT, SSEP, SCOM
STRANS, CRICSOL

FORM T₁, T₂, T₃ (TRANSFER FUNCTION PARAMETRIZATION)

Name: TFORM

Other modules required:

SMULT, SADD, DCOFAC

SPECTRAL FACTORIZATION

Name: SPECFAC

Other modules required:

ASDECOMP, SADD, SMULT, SINV
STRANS, CRICSOL, MSQRT, SSEP
SCOM, SHTRANS, MINIMAL

CO-SPECTRAL FACTORIZATION

Name: COSPECFAC

Other modules required:

STRANS, SPECFAC

INNER-OUTER FACTORIZATION

Name: IOFAC

Other modules required:

SHTRANS, SPECFAC, SMULT, SINV

CO-INNER-OUTER FACTORIZATION

Name: COIOFAC

Other modules required:

STRANS, IOFAC

COMPUTING THE DISTANCE FROM A GIVEN ANTISTABLE SYSTEM TO THE NEAREST STABLE SYSTEM

Name: HNORM

Other modules required:

SSEP, CLIASOL, SCOM, RSCHUR
SMULT

The reader at this point can appreciate the complexity of the software required to undertake an H_∞ design. The author believes that the module based approach is necessary to minimize development time and to maintain the required computer resources at a reasonable level.

SUGGESTIONS FOR FURTHER INVESTIGATION

It has already been noted that a complete evaluation of the applicability of H_∞ control to LSS problems must involve design studies, simulation studies, and ultimately tests on the MSFC LSS/GTF. To these ends, the obvious next step in the evaluation is to complete the development of a workable software system.

Additionally, there are several other areas of investigation that deserve attention.

Although the strategy of casting disturbance attenuation, tracking and robust stabilization problems via the standard problem formulation is very powerful when combined with H_∞ optimization, the actual design difficulties in LSS control involve the intelligent trade-off of performance and robustness. There are at least two approaches that immediately present themselves for further investigation. The first is to attempt to cast both of these goals simultaneously into the standard problem with a scalar "emphasis ratio" as the trade-off design parameter. The second approach is to borrow from the classical concept of minor loop design. In this approach, an inner loop design would first be accomplished with robust stabilization as the goal. An outer loop would then be designed with performance in mind. The idea is that the outer loop could be designed with a high degree of confidence in the model used, since the inner, or minor, loop had been designed to minimize the effects of plant variations.

Another issue of critical interest is that of controller complexity. H_∞ optimization techniques typically result in high order controllers. Preliminary results by Doyle et al. [20] indicate that the controller order can be maintained at the order of the open loop plant design model. However, this is probably unrealistic for the very high order models typical of LSS. A possible method for obtaining reduced order controllers is by utilizing balanced model reduction [21] on controllers obtained using the full design model. Glover [22] has shown that the degree of order reduction is closely related to closed loop stability and has derived an infinity norm bound that quantifies the allowable reduction. Safanov et al. [23] have recently presented reliable algorithms for calculating balanced reduced order controllers. It remains to be seen whether significant order reductions can be achieved in practical LSS problems.

Still another issue is whether the additional design complexity of H_∞ optimization gains enough in terms of the design goals to justify its use. The possibility of using the controller/transfer function parametrization to obtain non-optimal, but satisfactory, controllers has not been addressed.

Since the development of H_∞ optimization techniques is still in a state of evolution, it is difficult to make even a preliminary statement at this time regarding the eventual applications. There is no doubt, however, that the theory and computational techniques are sufficiently mature to begin an evaluation process.

REFERENCES

- [1] G. Zames, "Feedback and Optimal Sensitivity: Model Reference Transformations, Multiplicative Seminorms, and Approximate Inverses," IEEE Trans. Auto. Cont., vol. AC-26, pp. 301-320.
- [2] H. Kwakernaak and R. Sivan, Linear Optimal Control Systems, Wiley, New York, 1972.
- [3] M.G. Safonov and M. Athans, "Gain and Phase Margin of Multiloop LQG Regulators," IEEE Trans. Auto. Cont., vol. AC-22, pp. 173-178.
- [4] J.C. Doyle, "Guaranteed Margins for LQG Regulators," IEEE Trans. Auto. Cont., April 1978, pp. 756-757.
- [5] B.A. Francis, J.W. Helton, and G. Zames, " H_∞ Optimal Feedback Controllers for Linear Multivariable Systems," IEEE Trans. Auto. Cont., vol. AC-29, pp. 888-900.
- [6] L. Cheng and J.B. Pearson, "Frequency Domain Synthesis of Multivariable Linear Regulators," IEEE Trans. Auto. Cont., vol. AC-23, pp. 3-15.
- [7] C.A. Desoer, R.W. Liu, J. Murray, and R. Saeks, "Feedback System Design: the Fractional Representation Approach," IEEE Trans. Auto. Cont., vol. AC-25, pp. 399-412.
- [8] M. Vidyasagar, Control System Synthesis: A Factorization Approach, MIT Press, Cambridge, 1985.
- [9] R.D. Irwin, Lecture Notes on H_∞ Optimal Control Theory, given at NASA Marshall Space Flight Center, 1988.
- [10] B.A. Francis, "A Course in H_∞ Control Theory," Lecture Notes in Control and Information Sciences vol. 88, Springer-Verlag, 1987.
- [11] C.N. Nett, C.A. Jacobson, and M.J. Balas, "A Connection Between State Space and Doubly Coprime Fractional Representations," IEEE Trans. Auto. Cont., vol. AC-29, pp. 831-832.
- [12] M.G. Safonov and M.S. Verma, " L_∞ Sensitivity Optimization and Hankel Approximation," IEEE Trans. Auto. Cont., vol. AC-30, pp. 279-280.
- [13] H. Kwakernaak, "A Polynomial Approach to Minimax Frequency Domain Optimization of Multivariable Systems," Int. J. Control, vol. 44, pp. 117-156.

- [14] M.J. Grimble, "Optimal H_∞ Multivariable Robust Controllers and the Relationship to LQG Design Problems," Int. J. Control, vol. 48, pp. 33-58.
- [15] A.J. Laub, "A Schur Method for Solving Algebraic Riccati Equation," IEEE Trans. Auto. Cont., vol. AC-24, pp. 913-921.
- [16] G.H. Golub and C. Van Loan, Matrix Computations, Johns Hopkins Univ. Press, Baltimore, 1983.
- [17] D.Q. Mayne, "An Elementary Derivation of Rosenbrock's Minimal Realization Algorithm," IEEE Trans. Auto. Cont., vol. AC-18, pp. 306-307.
- [18] E.J. Davison, W. Gesing, and S.H. Wang, "An Algorithm for Obtaining the Minimal Realization of a Linear Time Invariant System and Determining if a System is Stabilizable-Detectable," IEEE Trans. Auto. Cont., vol. AC-23, pp. 1048-1054.
- [19] A.J. Laub, "Efficient Multivariable Frequency Response Computations," IEEE Trans. Auto. Cont., vol. AC-26, pp. 407-408.
- [20] J.C. Doyle, K. Glover, P. Khargonekar, and B. Francis, "State-Space Solutions to Standard H_2 and H_∞ Control Problems," Proc. Amer. Cont. Conf., 1988.
- [21] B.C. Moore, "Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction," IEEE Trans. Auto. Cont., vol. AC-26, pp. 17-31.
- [22] K. Glover, "All Optimal Hankel Norm Approximations of Linear Multivariable Systems, and Their L_∞ Error Bounds," Int. J. Control, vol. 39, pp. 1145-1193.
- [23] M.G. Safonov and R.Y. Chiang, "A Schur Method for Balanced Model Reduction," Proc. Amer. Cont. Conf., 1988.

N89-21744

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

The Measurement of Angular Differential Cross
Sections at the SSL Atomic Scattering Facility

Prepared by:	Thomas J. Kvale
Academic Rank:	Assistant Professor
University and Department:	The University of Toledo Department of Physics and Astronomy
NASA/MSFC:	
Laboratory:	Space Sciences
Division:	Solar-Terrestrial
Branch:	Atomic Physics
MSFC Colleague:	Dr. Marsha R. Torr
Date:	August 12, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

The Measurement of Angular Differential Cross
Sections at the SSL Atomic Scattering Facility

by

Thomas J. Kvale
Assistant Professor of Physics
The University of Toledo
Toledo, OH 43606 USA

ABSTRACT

This report covers the design of the SSL Atomic Scattering Facility (ASF) located at NASA/Marshall Space Flight Center as well as some of the initial experiments to be performed with it. The goal is to develop an apparatus capable of measuring angular differential cross sections (ADCS) for the scattering of 2 - 14 eV atomic oxygen from various gaseous targets. At present little is known about atomic oxygen scattering with kinetic energies of a few eV. This apparatus is designed to increase our understanding of collisions in this energy region.

Atomic oxygen scattering processes are of vital interest to NASA because the space shuttle as well as other low earth orbit satellites will be subjected to a flux of 5 eV atomic oxygen on the ram surfaces while in orbit. The primary experiments will involve the measurements of ADCS for atomic oxygen scattering from gaseous targets (in particular, molecular nitrogen). These, as well as the related initial experiments involving thermal He scattering from N_2 and O_2 targets will be described.

ACKNOWLEDGEMENTS

I would like to extend my thanks and appreciation to Dr. Marsha Torr for her support and enthusiasm for this project. Thanks and recognition also go to Dr. Charles Keffer and Dr. Glynn Germany for their guidance and discussions about the project. The majority of the design of the apparatus is due to their efforts.

I have enjoyed being part of the 1988 Summer Faculty Program and warm thanks are extended to Dr. Michael Freeman and Ms. Ernestine Cothran for their supervision. Finally, special thanks are extended to my wife, Barbara, and sons for their understanding.

INTRODUCTION

The Atomic Scattering Facility (ASF) is being constructed to perform experiments that will give us a better understanding of the atomic and molecular processes occurring in the altitude region of 200 - 500 km. As is shown in Figure 1, the primary species at this altitude is neutral atomic oxygen. Lesser amounts of N_2 , O_2 , and He are also present. Because the shuttle and other satellites operate in low earth orbit (LEO), it is of timely importance to understand these atomic scattering systems. The orbital velocity of the spacecraft at this altitude is approximately 8 km/s and from this reference frame, the spacecraft would be experiencing a flux of neutral atomic oxygen with 5 eV kinetic energy impinging on the ram surfaces.

The experiments, for which the ASF is designed, pose a number of experimental and theoretical challenges. All of the collisional partners are multi-electron and with the exception of helium, are of open shell configuration. This adds to the theoretical complexity of understanding the collision. Experimentally, the production and detection of few eV atomic oxygen is non-trivial. Interpreting the scattering is complicated by the need for crossed beams as well as accounting for the possibility of the incident atomic oxygen beam being in a mixture of the ground state $O(^3P)$ and the long-lived metastable state $O(^1D)$. However, when fully analyzed, angular differential cross sections (ADCS) have been used to understand more fully the potential energy surface governing the scattering. The acquisition of these data, for which this facility is designed, promises to have wide-ranging usefulness from both basic physics as well as engineering standpoints.

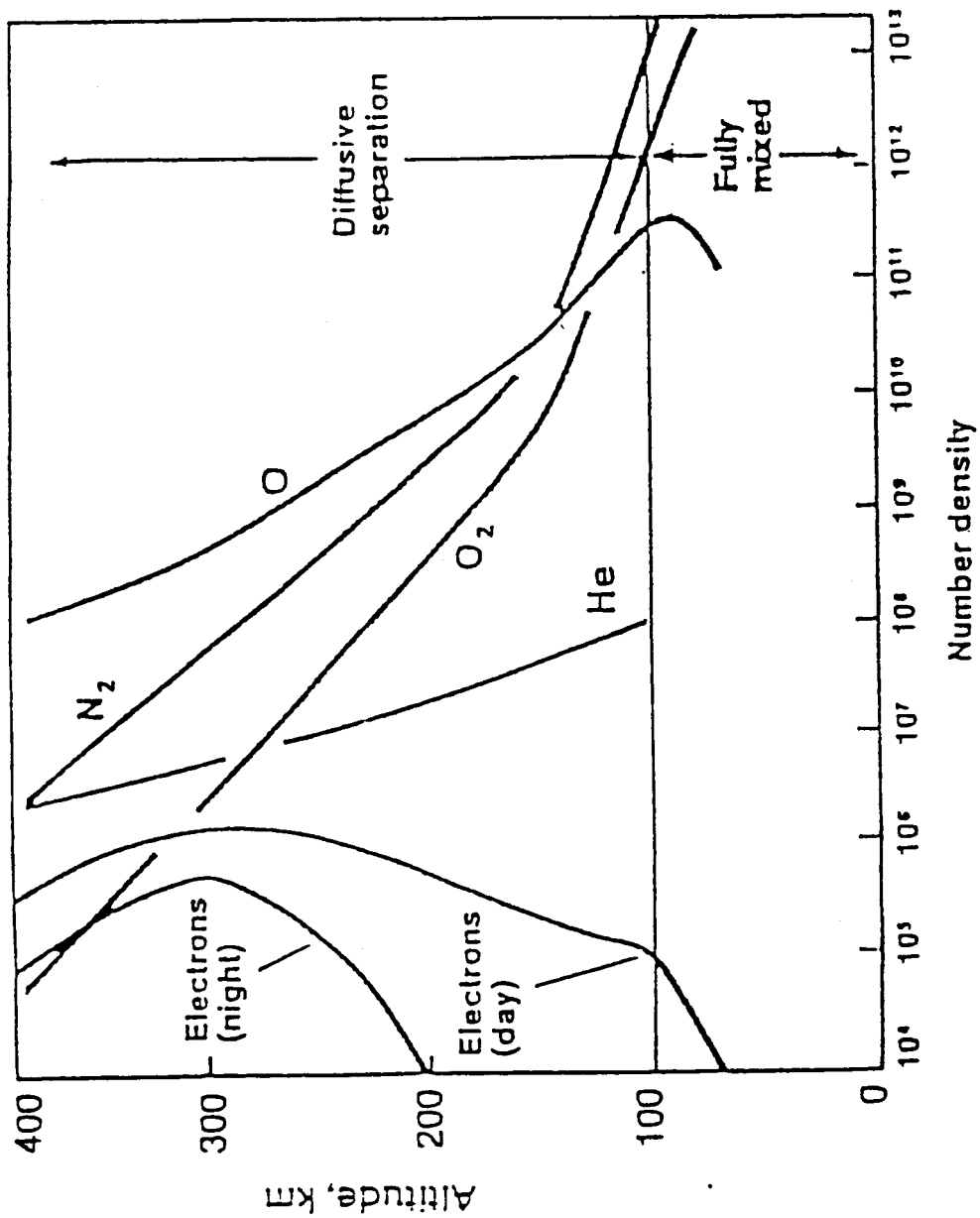


Figure 1. The number density (number/cm³) of several of the major species of elements constituting the atmosphere as a function of altitude. Low earth orbit (LEO) where the shuttle typically operates, is in the region of 300 km. The graph is from Ref. 1.

OBJECTIVES

This summer, I was involved in the following tasks:

- 1) To assist in the general design and construction of the Atomic Scattering Facility.
- 2) To outline the measureable variables necessary for the acquisition and analysis of angular differential cross section (ADCS) measurements.
- 3) To explore the possibility of normalizing the relative ADCS to an absolute measurement or calculation.

APPARATUS

The Atomic Scattering Facility (ASF) is designed to be able to obtain angular differential cross sections (ADCS) for neutral atom (or molecule) scattering from thermal neutral atoms or molecules. A schematic overview of ASF is shown in Figure 2, whereas a side view illustrating the placement of the major pieces of the apparatus is given in Figure 3. Ultra-high vacuum techniques have been applied in the general design of the facility. For example, the main vacuum chamber as well as most of the components are fabricated from 304 stainless steel. The projectile source, target jet, main chamber, and mass spectrometer are all differentially pumped to isolate the effects that gas loads will have on the other components. The base pressure in the main chamber is estimated to remain in the 10^{-6} to 10^{-7} Torr region while both the projectile source and target jet are operating, whereas the base pressure in the mass spectrometer will be in the 10^{-8} to 10^{-9} Torr region.

The initial major experiment guiding the construction of this apparatus is the measurement of ADCS for 5 eV atomic oxygen scattering from thermal molecular nitrogen. One of the distinguishing (and challenging) features of this apparatus is the production of the 5 eV atomic oxygen beam. The experimental arrangement that was chosen for this source of "fast" atomic oxygen follows a design² by Physical Sciences, Inc.³ Essentially, molecular oxygen is admitted into an expansion cone, as is shown in Figure 4, by a high-speed, pulsed nozzle valve. When a sufficient amount of molecular oxygen has been released to fill the space between the valve and the skimmer, a laser pulse from a 15 Joule CO_2 laser is focused at the throat of the nozzle. During the pulse, the thermal expansion of the gas accelerates the recombining atomic oxygen to eV kinetic energies. By varying the laser power and oxygen densities, the kinetic energy of atomic oxygen can be varied in the region of 2 - 14 eV. The E_{FWHM} of the atomic oxygen beam is estimated to be approx. 2 eV for a 5 eV nominal kinetic energy beam. The estimated incident flux directed on the target is 10^{12} to 10^{13} atoms/second.

The principal method of producing the N_2 target is by directing a thermal beam of molecular nitrogen into the center of the main vacuum chamber by a high speed, pulsed nozzle valve similar to that used with the oxygen source. The atomic oxygen beam and the thermal molecular nitrogen

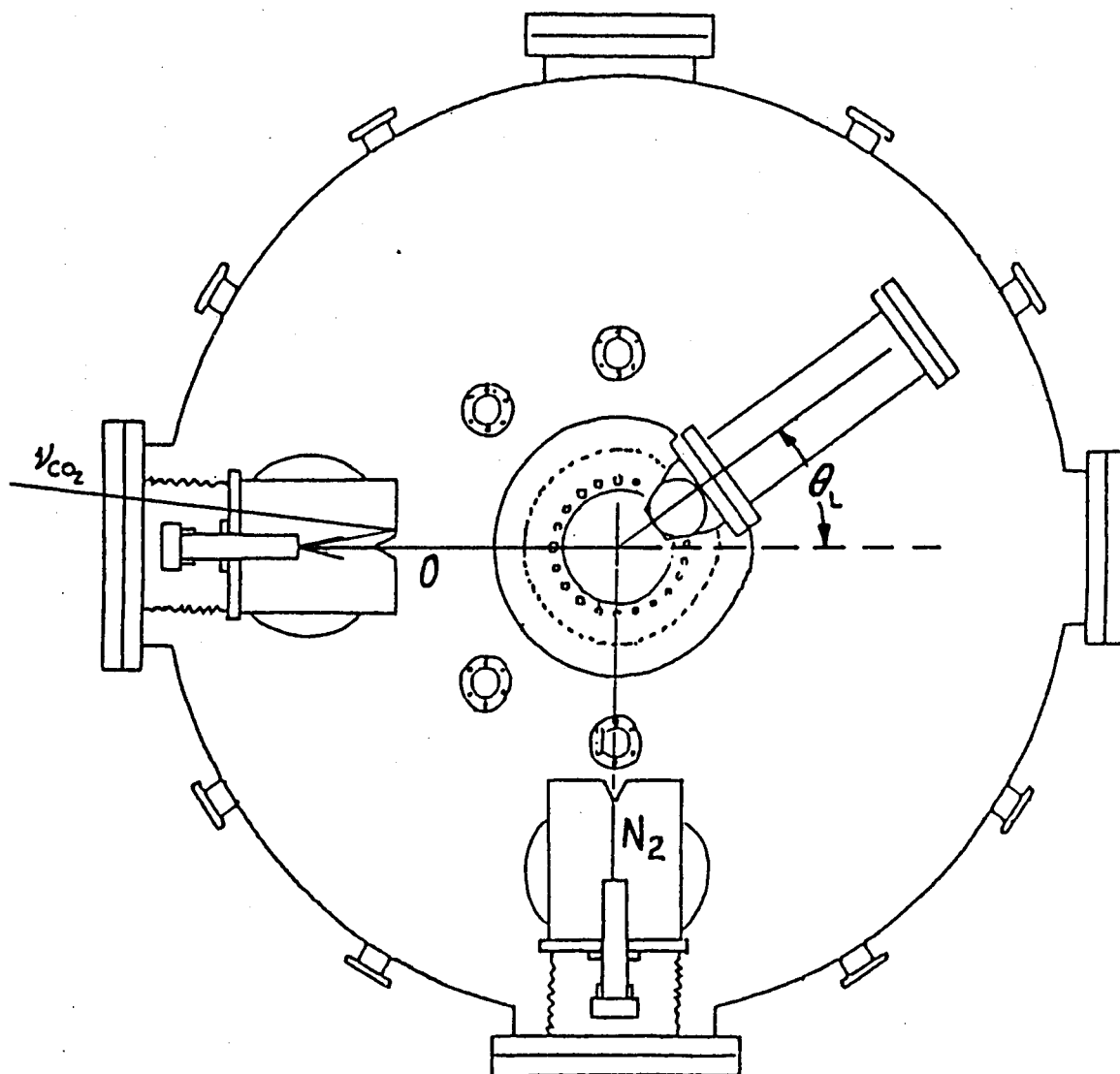


Figure 2. A schematic drawing of the overview of the experimental arrangement. Fast atomic oxygen is produced by a laser-generated plasma. The target species is directed to collide with the oxygen beam at the center of rotation of the mass spectrometer. The angle that the mass spectrometer makes with the oxygen beam is defined as the laboratory scattering angle, θ_L .

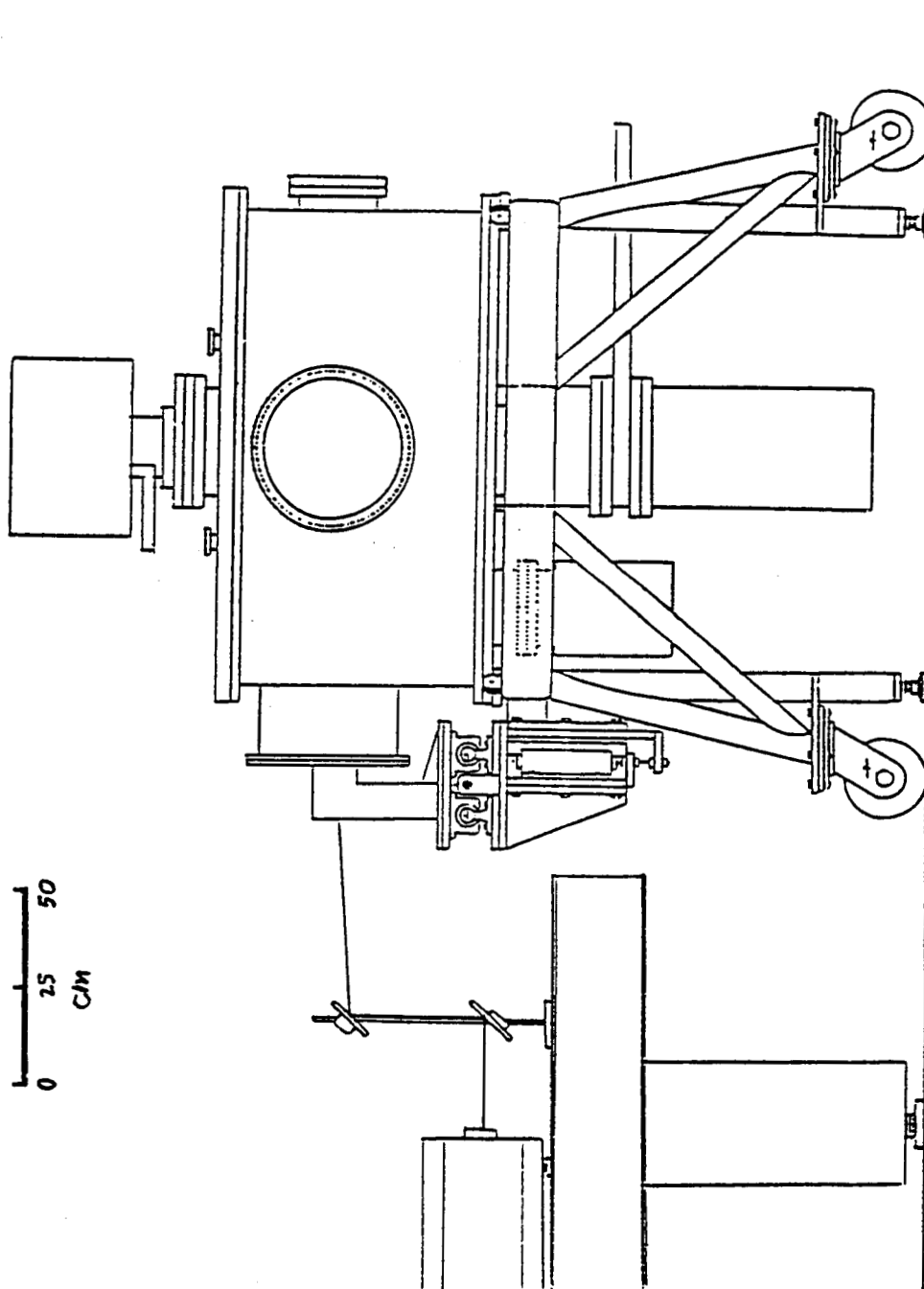


Figure 3. A schematic drawing of the major components of the SSL Atomic Scattering Facility. The CO_2 laser is located on an optical bench next to the main vacuum chamber. The scale gives an indication of the size of the facility.

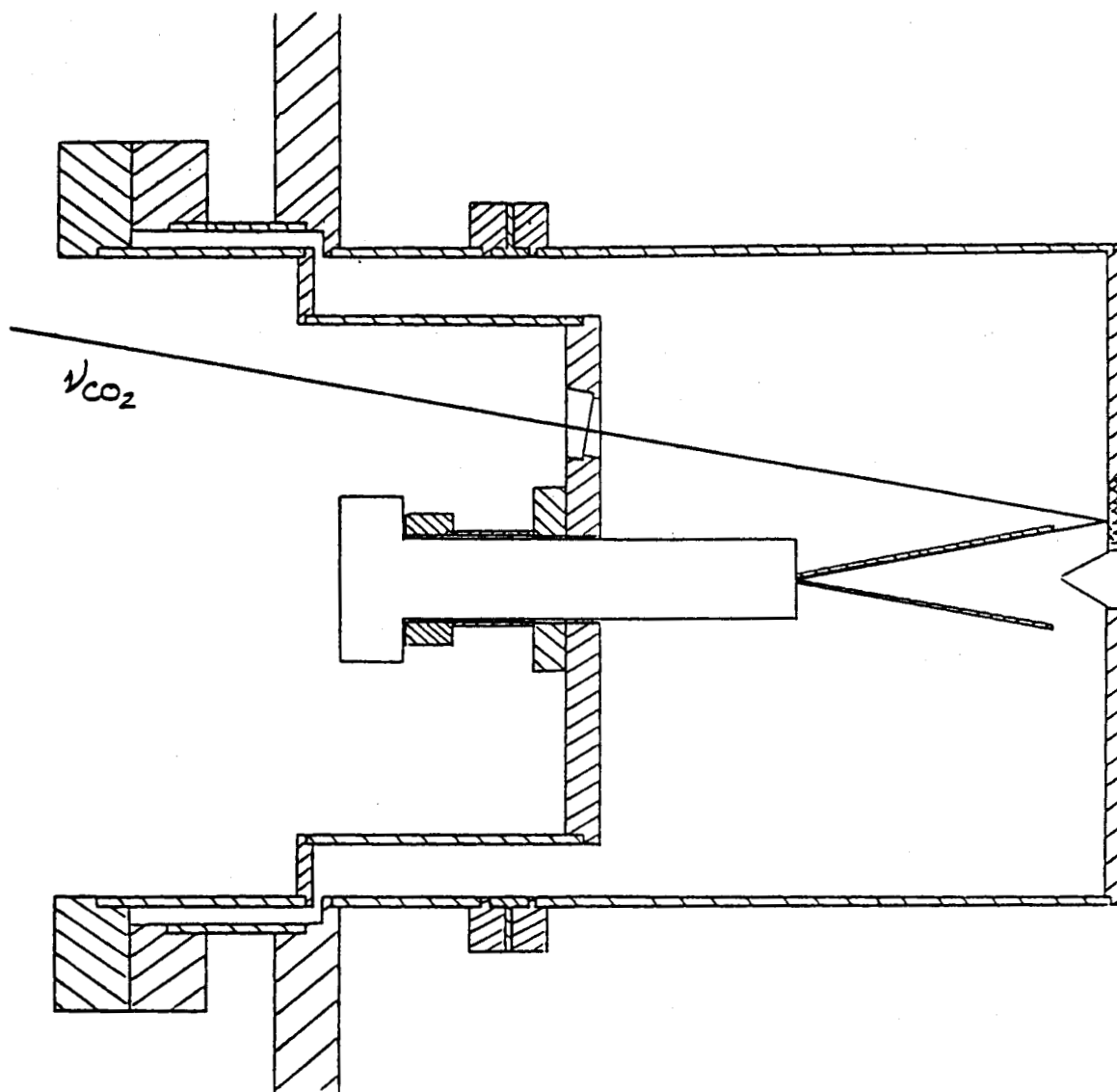


Figure 4. The schematic design of the fast atomic oxygen source. The laser light enters the differentially pumped housing through a vacuum window and is reflected by a mirror into the throat of the pulsed nozzle valve. The skimmer ensures a highly collimated beam.

beam intersect at right angles at the center of rotation of the mass spectrometer. This orientation reduces the effects that the spreads in the kinetic energies of the two beams have on the possible angular structure in the differential cross sections.

The scattered particles are detected by a mass spectrometer manufactured by Extrel Corp.⁴ It is differentially pumped by an ion pump, thereby maintaining a vacuum of approximately 10^{-9} Torr. The angle at which the mass spectrometer is located relative to the incident atomic oxygen beam is defined as the laboratory scattering angle. Rotation of the mass spectrometer is made possible by its mounting on a high vacuum rotation stage. The plane of rotation is the plane formed by the two intersecting particle beams, and the angular range of the system is limited only by the differential pumping housings of the devices. This ranges from 0 - 60 degrees on one side to 0 - 140 degrees on the other side. As a further check for the internal consistency of the data, the mass spectrometer assembly can be mounted on a port on the side of the main vacuum chamber and rotated perpendicular to the scattering plane. In this case, the angular range will be limited to only small angles.

The data acquisition electronics are shown in Figure 5, and the timing sequences for the various instruments are shown in Figure 6. In Figure 6, the baseline for each instrument represents an "off" state and raised step represents the "on" or activated state. The pulsed operation is necessary due to the use of a pulsed laser in the production of atomic oxygen and helps reduce the gas load on the system. The timing of the various components is dictated by the kinematics of that particular experiment. The O_2 pulse length is related to the time required to fill the cone region shown in Figure 4, whereas the target pulse length is to produce quasi-static number density conditions when the atomic oxygen beam arrives at the interaction region. The target is assumed to have the lesser velocity, so its pulse will initiate a data sequence. The O_2 pulse is delayed ($t_0 = 350 \mu s$) to allow for the different velocities of the two beams and the laser delay ($t_1 = 100 \mu s$) is to allow for the expansion of O_2 into the cone region. The counter is gated to help reduce the background signals.

Bailey and Mullen⁵ have examined the effects that thermal motion of the target have on the observed ADCS. In their study, the target configuration was varied to be: a room-temperature (300 K) static gas cell, a static gas cell at 77 K, or a beam intersecting the projectile beam at 90

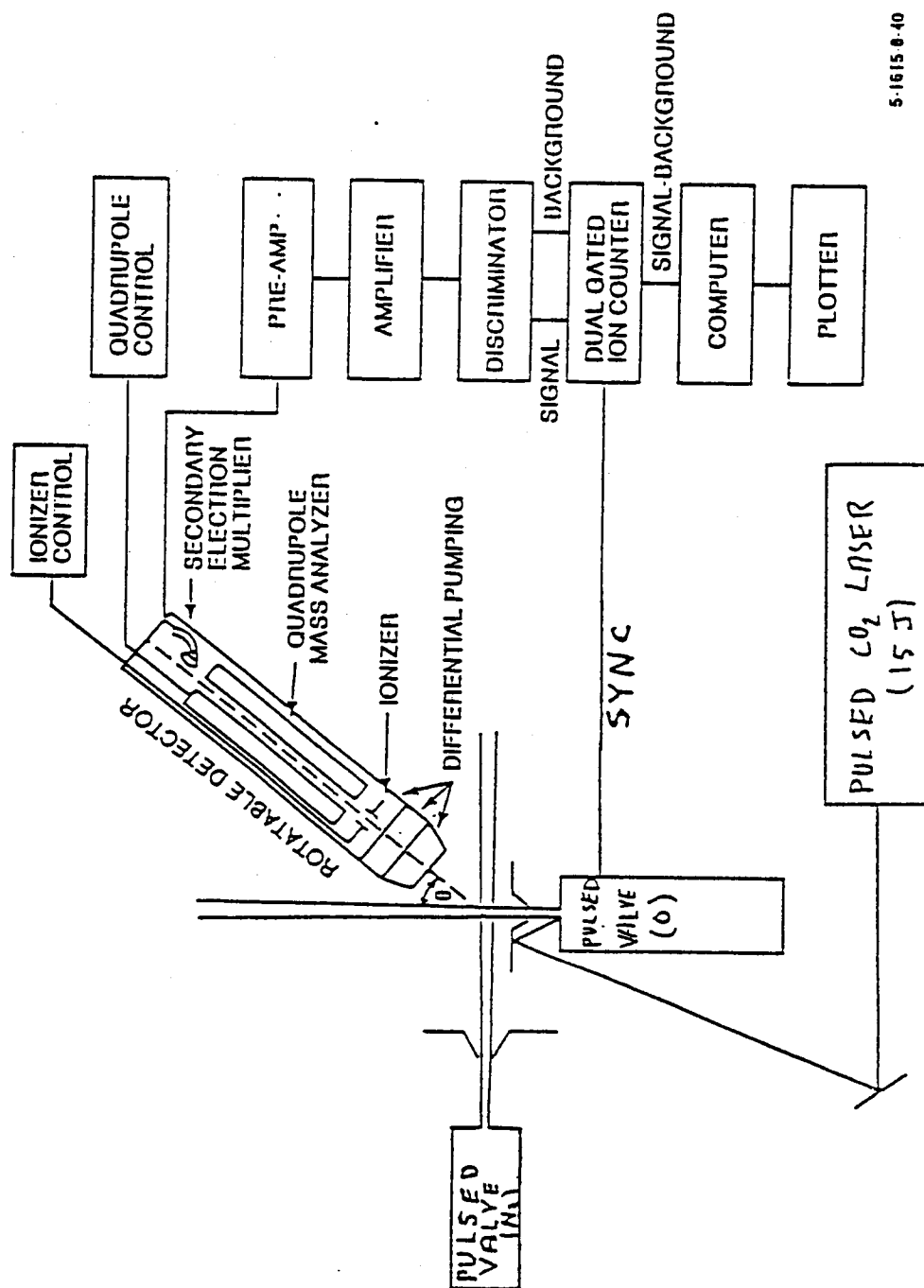


Figure 5. A schematic drawing of the electronics necessary for the data acquisition system.

TIMING DIAGRAM

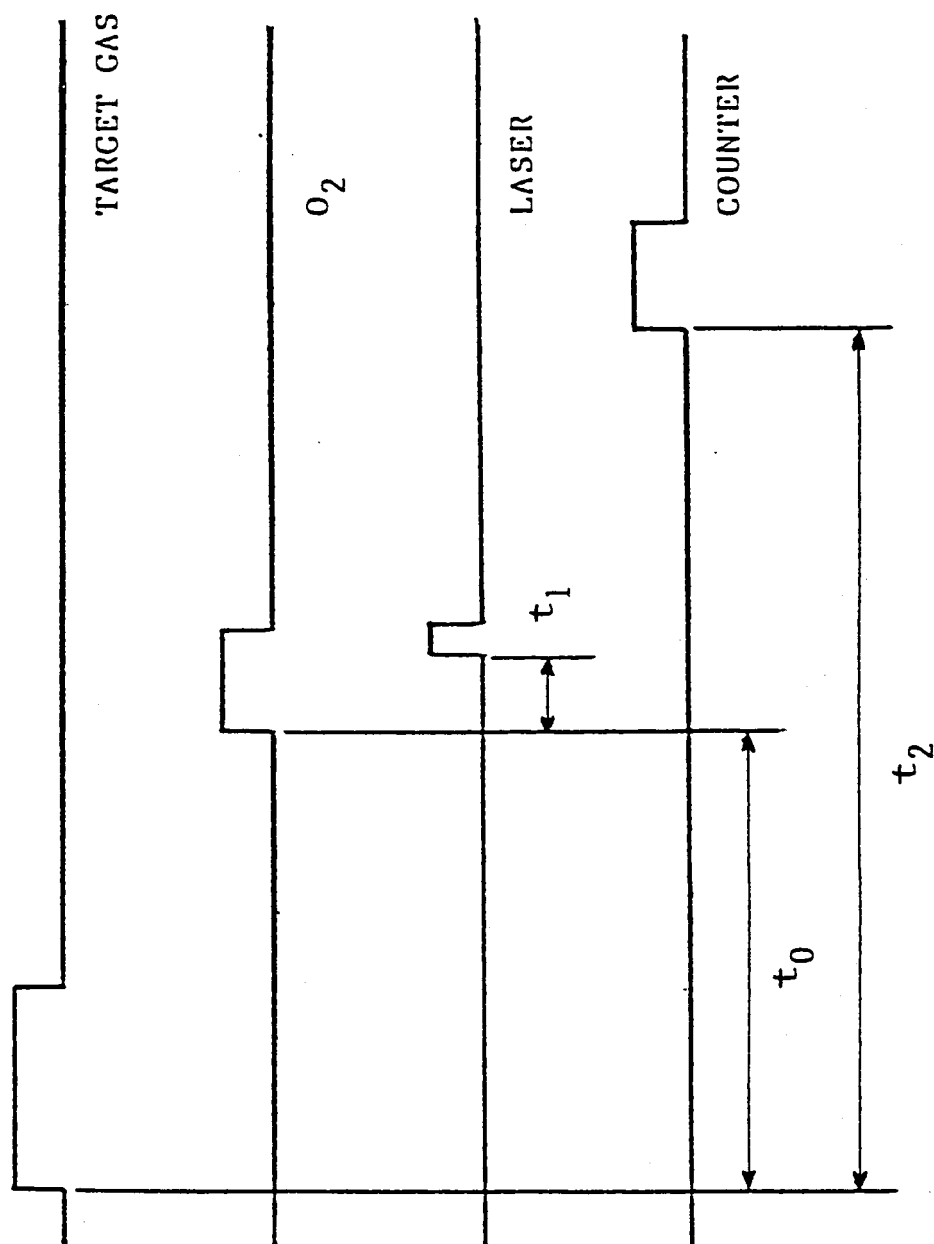


Figure 6. The timing diagram for the various events necessary for the ADCS measurements.

degrees. Their results indicated that the room-temperature gas cell had the greatest effect on averaging over possible angular structure in the ADCS, whereas the crossed beam configuration had the least effect. Because of this, it was decided to use the crossed beam configuration for the majority of the measurements and possibly use the static gas cell only for normalization of the relative measurements.

DIFFERENTIAL CROSS SECTION MEASUREMENTS

There are several different ways to express the variables necessary for angular differential cross section (ADCS) measurements. Following the notation of Park and coworkers,⁶⁻⁸ an apparent differential cross section is given by:

$$\frac{ds(\theta, E)}{d\Omega} = \frac{I(\theta)}{I_0 n l \Delta\Omega} \quad (1)$$

where $I(\theta)$ is the scattered signal measured at the lab scattering angle θ_L , I_0 is the total integrated current (see the discussion in Refs. 6-8 whether this should be incident or elastic), n is the target number density, l is the interaction path length, and $\Delta\Omega$ is the solid angle subtended by the detector. In our case, the target beam is produced by a free jet expansion technique and thus $n l$ is not known precisely. Because of this, the ADCS will be reported as relative measurements. A scale factor can, at a later time, normalize these relative measurements to produce absolute ADCS. Because both the source and target are crossed beams, and the mass spectrometer rotates in the plane formed by these two beams, $n l$ will also have an angular dependency that must be accounted for. The term "apparent" ADCS ($ds/d\Omega$) is used in Equation 1 because several apparatus effects must be removed from the data before "true" angular differential cross sections ($d\sigma/d\Omega$) can be obtained. There are several methods that have been employed to arrive at $d\sigma/d\Omega$, and the reader is referred to one of the references.⁶⁻¹⁰ For the remainder of this report, it is assumed that "true" ADCSs (either relative or absolute) have been obtained unless stated otherwise.

The kinetic energy of the scattered projectile will change depending on the angle of deflection it occurs during the collision. This kinematic energy loss is the result of the transfer of kinetic energy from the projectile to the recoiling target. Table I shows the various energies and angles in the laboratory reference frame of the collisional partners. Because the detection efficiency of the mass spectrometer is energy dependent and the scattered projectile energy changes significantly as a function of scattering angle, the data will have to be adjusted to account for this variation. A preliminary analysis of this kinetic energy loss was conducted by the approximation that the thermal N_2 beam is at rest. This approximation was used in the data shown in Table I. A more realistic model will

Table I

ORIGINAL PAGE IS
OF POOR QUALITY

Kinematic Scattering Program — T. J. Kvale — 09 August 1988
 Eprojectile(eV) = 5.0000 Etarget(eV) = 0.0000
 Mprojectile(amu) = 16.00 Mtarget(amu) = 28.00
 Target Angle(deg) = 0.000
 Internal state excitation energy (eV) = 0.0000

Scattering Angle(deg)	Escat(eV)	Recoil Angle(deg)	Erecoil(eV)
0.000	5.0000	180.000	0.0000
5.000	4.9783	273.927	0.0217
10.000	4.9137	277.847	0.0863
15.000	4.8080	281.753	0.1920
20.000	4.6638	285.635	0.3362
25.000	4.4849	289.487	0.5151
30.000	4.2759	293.301	0.7241
35.000	4.0418	297.066	0.9582
40.000	3.7884	300.775	1.2116
45.000	3.5216	304.416	1.4784
50.000	3.2473	307.980	1.7527
55.000	2.9716	311.455	2.0284
60.000	2.6996	314.831	2.3004
65.000	2.4364	318.095	2.5636
70.000	2.1860	321.239	2.8140
75.000	1.9516	324.251	3.0484
80.000	1.7357	327.123	3.2643
85.000	1.5395	329.849	3.4605
90.000	1.3636	332.425	3.6364
95.000	1.2079	334.849	3.7921
100.000	1.0713	337.123	3.9287
105.000	0.9528	339.251	4.0472
110.000	0.8506	341.239	4.1494
115.000	0.7632	343.095	4.2368
120.000	0.6888	344.831	4.3112
125.000	0.6258	346.455	4.3742
130.000	0.5726	347.980	4.4274
135.000	0.5280	349.416	4.4720
140.000	0.4908	350.775	4.5092
145.000	0.4601	352.066	4.5399
150.000	0.4349	353.301	4.5651
155.000	0.4146	354.487	4.5854
160.000	0.3987	355.635	4.6013
165.000	0.3868	356.753	4.6132
170.000	0.3784	357.847	4.6216
175.000	0.3735	358.927	4.6265
180.000	0.3719	360.000	4.6281

account for the thermal velocity of the target. Work is in progress on the modifications to the computer code to incorporate this condition.

There are a variety of methods that have been employed to obtain absolute angular differential cross sections (ADCS). Included in this report is only a brief introduction and the reader is referred to the references for more complete descriptions. An exhaustive review of all the work devoted to it is beyond the scope of this report.

- 1) The direct measurement of all parameters required for ADCS.^{6-8,10}

This requires that all the parameters in Eq. 1 are known exactly. This has been applied to experiments in the keV impact energy region with success. In these experiments, the scattered beam is treated identical to the incident beam, thereby cancelling the effect that the detector efficiency has on the measured currents $I(\theta)$ and I_0 in Eq. 1. Also, n , l , and $\Delta\Omega$ are geometrically defined or measured. In the proposed experiment, these measurements would be difficult and accounting for the kinematic energy loss by an increase in the projectile energy would vary significantly that energy.

2. Normalize to previous absolute work.¹¹⁻¹³

Essentially, this would place a multiplicative factor in Eq. 1 and would account for the uncertainty in the exact value of n . Experimentally, the relative ADCS could be normalized to other (or the same) targets under similar experimental conditions. A second option would be to integrate the relative ADCS to obtain total cross sections (TCS) and then normalize the TCS to absolute TCS. The third method of normalization involves normalizing the relative ADCS (or TCS) to a reliable theoretical calculation. The fact that the atomic oxygen source may produce a mixture of $O(^3P)$ and $O(^1D)$ states, as well as the inability of this apparatus to distinguish ro-vibrationally inelastic scattering complicates this last approach.

RELATED WORK

Although ADCS for 5 eV atomic oxygen scattering from molecular nitrogen have not been reported, an indication for it can be obtained from the keV scattering data using the classical approximation.^{10,14,15} The basic tenets of this approximation involve the role that the impact parameter b plays in the scattering. For small angle scattering, the product of impact energy times the scattering angle can be expanded in a series in b ,

$$\chi(E\theta) = E\theta = \chi_0(b) + E^{-1}\chi_1(b) + \dots \quad (2)$$

The reduced differential cross section can also be expanded in a series of χ ,

$$\rho(\theta, E) = \theta \sin \theta \frac{d\sigma}{d\Omega} = \rho_0(\chi) + E^{-1}\rho_1(\chi) + \dots \quad (3)$$

If it is valid to truncate both series with the first term, then it is a straightforward task to use the higher energy keV data to predict what the 5 eV ADCS will look like.¹⁵⁻¹⁷ This approach was taken by K. Smith¹⁷ and his results are shown in Figure 7.

A preliminary series of experiments can be performed prior to the $O + N_2$ experiment which use thermal helium as the projectile rather than atomic oxygen. This will provide the opportunity to compare with previous measurements. The work by Keil et al.¹⁸ have the possibility of detecting scattering both in the scattering plane and a right angles to it. The work by Beneventi et al.¹⁹ is a later, high resolution study, and their results are shown in Figure 8. This series of experiments will require only minimal modifications to the experimental procedure and will be useful both as a check against systematic errors and as an extension of their work.

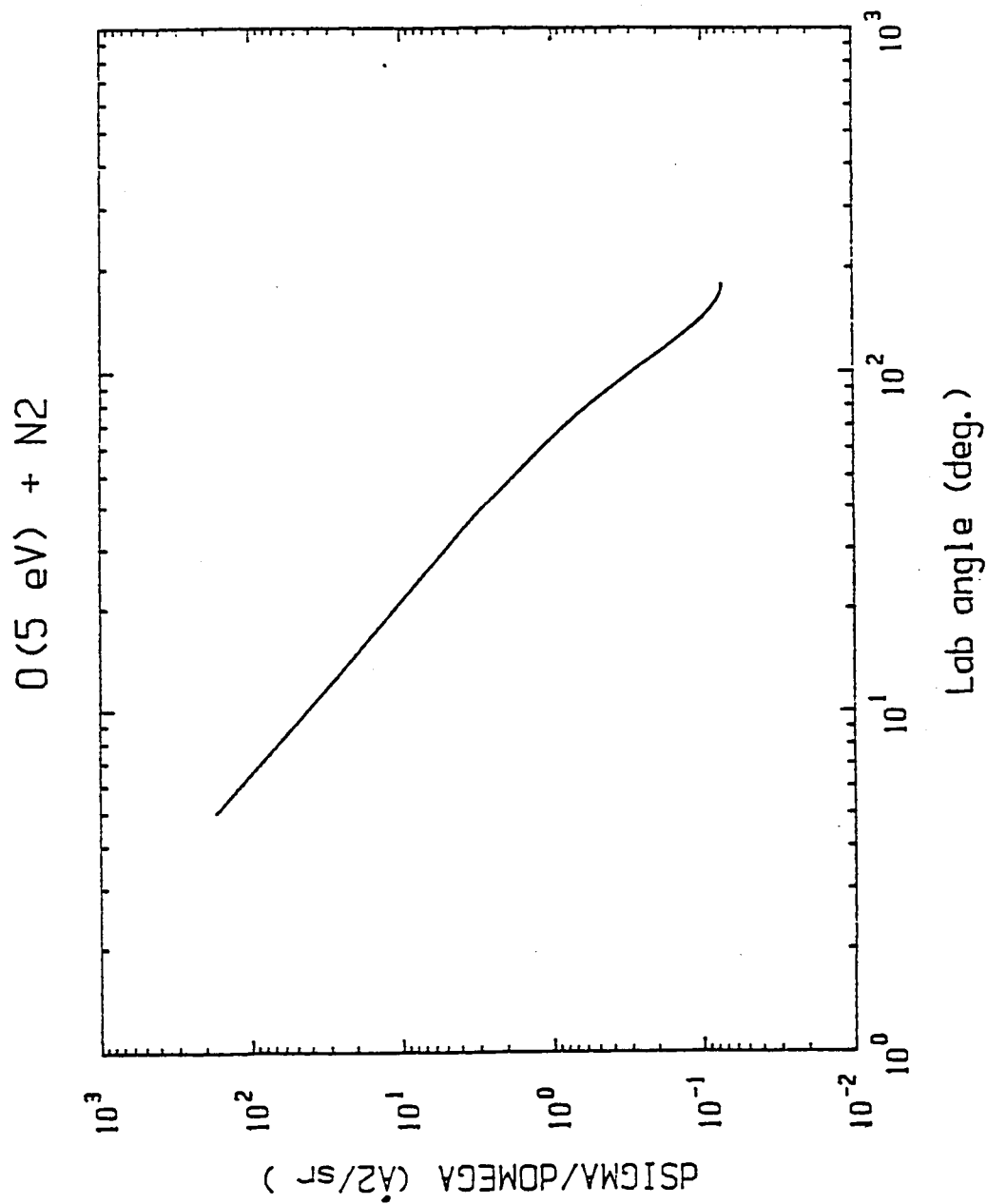


Figure 7. The prediction of $O(5 \text{ eV}) + N_2$ angular differential cross sections from Ref. 17.

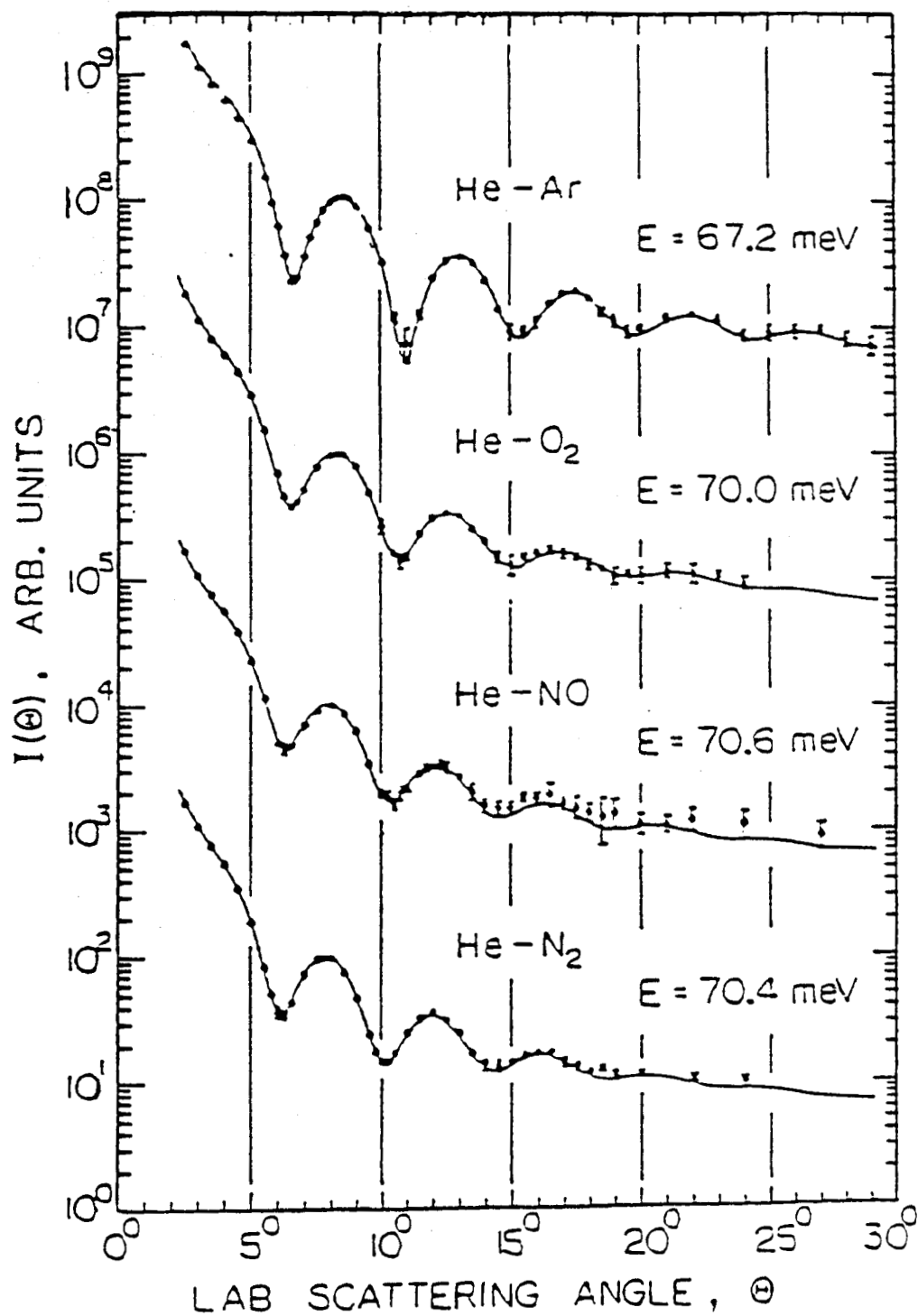


Figure 8. Thermal energy helium scattering from various targets. Graph is from Ref. 18.

CONCLUSION AND RECOMMENDATIONS

This report contained only a brief introduction to the SSL Atomic Scattering Facility (ASF). The design of it is in an advanced stage, and components are being acquired. The parameters required for the measurement of angular differential cross sections (ADCS) have been outlined and their implementation in this apparatus was discussed. In addition, several methods that have been used on other experiments to arrive at absolute differential cross sections were listed. The angular dependence of the ADCS will give an indication as to which method would be best suited for this experiment.

Some of the experimental tasks include the preliminary thermal He + (N₂, O₂) relative ADCS measurements. This is useful to provide a check against systematic errors in the new facility. Two independent measurements by other groups will provide an opportunity for comparison. The output and operating variables of the atomic oxygen source need to be characterized. The large energy spread and the mixture of atomic oxygen states (³P and ¹D) present in the incident beam will complicate the understanding of the scattering. The functional shape of the relative ADCS will indicate the most desirable method of normalization. Once the preliminary and O + N₂ experiments have been completed, a whole range of other experiments involving other targets will be possible.

The SSL Atomic Scattering Facility is designed to address some uncertainties in our knowledge of few eV atomic oxygen scattering. This is a difficult impact energy region to study experimentally due to the complexity of producing intense few eV neutral atomic oxygen beams. Also, few theoretical calculations are available in this energy region for neutral atomic oxygen impact. For these reasons, this apparatus should make a significant contribution to the understanding of neutral atomic oxygen interaction with various atoms or molecules in the few eV impact energy region.

REFERENCES

1. R. G. Roble, Chem. Eng. News, p.23 (16 June 1986).
2. G. E. Caledonia and R. H. Krech, AIAA paper 85-7015, p. 153 (1985).
3. Physical Sciences, Inc., Andover, MA 01810.
4. Extrel Corp., Pittsburgh, PA 15238.
5. T. L. Bailey and J. M. Mullen, Int. J. Mass Spect. 18, 339 (1975).
6. J. T. Park, J. M. George, J. L. Peacher, and J. E. Aldag, Phys. Rev. A18, 48 (1978).
7. J. T. Park, IEEE Trans. Nucl. Sci. NS-26, 1011 (1979).
8. J. L. Peacher, T. J. Kvale, E. Redd, P. J. Martin, D. M. Blankenship, E. Rille, V. C. Sutcliffe, and J. T. Park, Phys. Rev. A26, 2476 (1982).
9. F. von Busch, J. Phys. B8, 1440 (1975).
10. D. C. Lorents and W. Aberth, Phys. Rev. 139, 1017 (1965).
11. D. F. Register, L. Vuskovic, and S. Trajmar, J. Phys. B19, 1685 (1986).
12. D. F. Register, H. Nishimura, and S. Trajmar, J. Phys. B13, 1651 (1980).
13. J. T. Park, J. E. Aldag, J. L. Peacher, and J. M. George, Phys. Rev. A21, 751 (1980).
14. F. T. Smith, R. P. Marchi, W. Aberth, D. C. Lorents, and O. Heinz, Phys. Rev. 161, 31 (1967).
15. Y. N. Belyaev and V. B. Leonas, ZhETF Pis'ma 4, 134 (1966).
16. D. A. Schafer, J. H. Newman, K. A. Smith, and R. F. Stebbings, J. Geophys. Res. 92, 6107 (1987).
17. K. Smith, private communication (1988).

18. M. Keil, J. T. Slanka, and A. Kuppermann, J. Chem. Phys. 70, 482 (1979); J. Chem. Phys. 70, 541 (1979).
19. L. Beneventi, P. Casavecchia, and G. G. Volpi, J. Chem. Phys. 85, 7011 (1986).

N89-21745-99-93

348.

187268

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

**MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA**

**Angular Response Calibration of the
Burst And Transient Source Experiment**

M3116546

Prepared by:

John Patrick Lestrade

Academic Rank:

Assistant Professor

University and Department:

Mississippi State University
Department of Physics and
Astronomy

NASA/MSFC:

Laboratory:

Space Science

Division:

Astrophysics

Branch:

High-Energy Astrophysics

MSFC Colleague:

G. J. Fishman

Date:

July 29, 1988

Contract No.:

NGT 01-002-099
University of Alabama

Angular Response Calibration of the Burst And Transient Source Experiment

by

John Patrick Lestrade
Assistant Professor of Physics and Astronomy
Mississippi State University
Mississippi State, Mississippi, 39762

Abstract

The Gamma Ray Observatory includes four experiments designed to observe the gamma-ray universe. Jerry Fishman in the High Energy Astrophysics Branch at Marshall Space Flight Center is the principal investigator for one of these experiments, the Burst And Transient Source Experiment (BATSE).

We have recently completed laboratory measurements to test the response of the BATSE modules to gamma-ray sources that are non-axial. The results of these observations are necessary for the correct interpretation of BATSE data obtained after it is put in Earth orbit. The launch is planned for March, 1990.

Preliminary analyses of these test data show the presence of a radial dependence† to the detector's light collection efficiency. As a continuation of the work begun as a NASA/ASEE Summer Faculty Fellow, we propose to evaluate the importance of this radial response, analyze future experimental data to derive the actual functional dependence on radius, and calculate the net effect on the output spectrum as a function of the angle of incidence.

† The dependence is radial in the sense that energy-depositing events that cause scintillation near the center of the 20-inch diameter disc are more efficiently collected than those events near the edge.

Acknowledgement:

This past summer has been an invaluable experience for me, both professionally and personally. The BATSE team have given unselfishly of their time and expertise to make me an integral part of the group. I appreciate all that I have learned from them. I am sincerely grateful for their acceptance and friendship.

In particular I would like to thank Jerry Fishman. His knowledge and enthusiasm have been very helpful. His guidance throughout this summer is responsible for the positive results that I have obtained.

This experience would not have been realized if it were not for the friendship and help of Ernestine Cothran and Michael Freeman. As coordinators of the program, they are responsible for my participation.

Pat Lestrade
July 29, 1988

List of Figures

	page
1. The Gamma-Ray Observatory (GRO)	XIX-14.
2. BATSE detector module	XIX-14.
3. A generic gamma-ray spectrum (Cesium 137)	XIX-15.
4. The August 5, 1985 gamma-ray burster event	XIX-16.
5. The spectrum of the March 7, 1979 gamma-ray burster event	XIX-16.
6. The effect of the slant path on efficiency	XIX-17.
7. The angular dependance of photopeak height (Cesium 662 keV)	XIX-18.
8. Angles of incidence of TPS-19	XIX-19.
9. Cesium spectra at three different angles for the Spectroscopy detector (SPD)	XIX-20.
10. Cesium spectra at four different angles for the Large Area detector (LAD)	XIX-21.
11. Angular dependence of photopeak centroid energy (Cs-137, Se-75, and Cd-109)	XIX-22.
12. Gauss fit to the Cs 662 keV peak at 0°	XIX-23.
13. Gauss fit to the Cs 662 keV peak at 266°	XIX-24.
14. Angular dependence of detector resolution (FWHM/Energy) for Cs 662 keV peak	XIX-25.
15. Angular dependence of photopeak height for four different energies	XIX-26.
16. Sodium-22 1275 keV peak as recorded by three different Large Area detectors	XIX-27.
17. Sodium-22 511 keV peak as recorded by three different Large Area detectors	XIX-28.
18. Energy vs. channel for the Large Area detector of module 1	XIX-29.
19. Energy vs. channel for the Spectroscopy detector of module 1	XIX-30.

1. Introduction:

1.1 The Gamma-Ray Observatory

Before the year 2000 NASA plans to launch more than 9 major space missions. These observatories will study the universe in the infrared, visible, ultraviolet, x-ray, and gamma-ray portions of the spectrum. The objects to be studied include the earth, the planets, the sun, and other more exotic, cosmological objects in which high-energy processes are taking place. Some of the objects will be well known, at least theoretically, such as pulsars. Others, for example, gamma-ray bursters, are mysterious and enigmatic.

One of these future satellites is the Gamma-Ray Observatory (GRO). Shown in Figure 1, GRO is a 17-ton satellite carrying 4 experiments; the Oriented Scintillation Spectrometer Experiment (OSSE), the Imaging Compton Telescope (COMPTEL), the Energetic Gamma-Ray Experiment (EGRET), and the Burst And Transient Source Experiment (BATSE). Together, these instruments detect gamma radiation at energies from 30 keV to 30 GeV. GRO is scheduled to be launched in March of 1990 to a nominal altitude of 250 miles (400 km). Its expected lifetime is from two to ten years.

The High-Energy Astrophysics Branch in the Space Science Laboratory at Marshall Space Flight Center has designed, built, and is now testing the BATSE detectors. As a NASA/ASEE Summer Faculty Fellow I have been involved in some of the laboratory testing of the BATSE modules and in the preliminary analysis of the acquired data.

1.2 Burst And Transient Source Experiment

BATSE is composed of eight detector modules (cf. Figure 2). These modules will be situated at the eight 'corners' of GRO. In this orientation it is able to monitor the complete sky except for that part temporarily blocked by the earth. As the name implies, BATSE is designed to observe and record gamma-ray events that are short-lived. The principal transient source to be studied will be gamma-ray bursters. At the same time, it has the ability to observe sources of gamma radiation that are long-lived – such as the sun, pulsars, and black holes.

The principal detector in each module, called the Large-Area detector (LAD), is a sodium iodide crystal in the shape of a disc with a diameter of 20 inches and a thickness of 0.5 inches. The shape and size of the crystal were chosen to make

BATSE more sensitive to the low-energy gamma ray spectrum (i.e., 30 keV~ 1.8 MeV) and to permit the measurement of very weak sources.

Also shown in Figure 2, in front of the LAD, is a charged-particle detector (CPD) consisting of a plastic scintillator. Charged particles such as cosmic rays that deposit energy in the LAD will also be recorded in the CPD. Electronic coincidence circuitry is used to exclude these events from the recorded spectra.

In addition to the main crystal, there is a secondary detector called the Spectroscopy detector (SPD). It too is composed of sodium iodide, 3-inches thick with a 5-inch diameter. Because of its greater thickness, it extends the sensitivity to higher energies (≈ 25 MeV). It also has significantly better resolution than does the LAD.

1.2.1 BATSE Testing and Calibration

One of the problems inherent in an observational science is the effect that the observer's instrument has on the data. All instruments have what is called a 'point-spread function'. That is, given a uniquely-valued input, the output contains a finite spread in values reflecting the instrument's resolution. In the case of a gamma-ray instrument, such as BATSE, there are contributions to this dispersion from inhomogeneities in the crystal, from statistical fluctuations in the conversion of gamma-ray energy into an electrical signal, from electronic noise, and from other sources. Each contribution tends to spread the monoenergetic input into an approximate gaussian shape. An example of the resultant 'photopeak' is shown in Figure 3. However, the results are further complicated by scattering events which result in only partial deposition in the crystal. The total 'spread' is much more complicated.

When a monoenergetic beam of photons strikes the sodium iodide crystal in the detector, some of the photons are completely converted into an electrical response that becomes recorded as counts in one of the channels of the photopeak (see Figure 3). However, some photons deposit only part of their energy and are scattered out of the detector. These become recorded as lower-energy events and show up as counts in the Compton continuum. Still other photons lose some energy outside of (often behind) the crystal and are then scattered back into the crystal where they become recorded as counts in the backscatter peak. As seen in this figure, even when the input is monoenergetic, these effects distort the spectrum in a very complicated way. When the source is *not* monoenergetic the situation is considerably more complicated.

Figures 4 and 5 present observations of two typical gamma-ray bursters†. The first, from a Spacelab 2 measurement on Aug. 5, 1985, shows that the duration of

† Harding *et. al.*

a burst is on the order of seconds. The next figure (5) shows the spectrum of a burster on Mar. 7, 1979. The exponential spectrum is typical of these events and is consistent with a model based on thermal bremsstrahlung. These are certainly more complicated than the monoenergetic case. Before we can determine the true spectrum that was *incident* on the detector, we must solve a complicated problem that depends on how the detector responds to incident photons. This response is a function of not only the incident energy but also of the angle of incidence. The task of developing this deconvolution software has been assigned to a BATSE co-investigator at the Goddard Space Flight Center. At MSFC part of our responsibility is to provide a response matrix that describes the detector's effects on an input spectrum. To this end we have spent the last few months measuring the spectra of several sources under various conditions. This report outlines some of those measurements and proposes some tests and procedures for the future.

2. Past Work

2.1 Angular Response: TPS-19

During a gamma-ray event for BATSE in earth orbit there are at least four detectors responding to the flux of radiation. Since these detectors are oriented in different directions, it is important to know how the response changes as a function of the angle of the incident flux.

The most important factor in this response is the projected area of the detector face to the radiation. This should, and does, impose a (co)sinusoidal response. However, there are several other significant factors that affect the response.

The first of these is an effect that counteracts the cosine decrease. At near-normal incidence ($\theta = 0^\circ$) as the angle increases there is an *increase* in the deposition probability for photons entering the 0.5-inch thick crystal due to the increasing slant path. This opposes the decrease in absorption caused by the decrease in projected area. To a first approximation the angular response near 0° is

$$\cos \theta \cdot \left(1 - e^{-\tau/\mu}\right) \quad (1)$$

where $\mu = \cos \theta$. For sodium iodide the absorption coefficient at 662 keV, is approximately 0.71 in^{-1} (0.28 cm^{-1}). This gives a τ of 0.36 for a half-inch normal thickness. This function (1) and the cosine function are plotted in Figure 6. Measurements of spectra to be shown below will verify this slant-path effect.

The second factor arises because the detector has a finite thickness. Therefore, even at an angle of incidence of 90° , there is some deposition of energy in the crystal. Figure 7 presents observations of the height of the cesium-137 662 keV photopeak as a function of angle. At 90° the data are non-zero.

The third effect is seen at angles greater than 90° . It is due to the complicated scattering and absorption introduced by the material behind the crystal. At high enough energies (i.e., $> 150 \text{ keV}$) some radiation is still deposited in the crystal. However, it is attenuated compared with that at 0° . A later figure covering 360° will show more structure because of the heterogeneous distribution of the electronics.

2.2. Experimental Considerations

Ideally, we would like to be able to irradiate the instrument with a monoenergetic beam of gamma rays with a constant flux. This would make the analyses

of the output spectra a simpler task. As mentioned above, however, even such a simple beam produces a complicated spectrum. In addition, this "response" will vary as the instrument is turned with respect to the incident beam. In order to get the perfect response matrix therefore, we would have to repeat those measurements for all angles and at all energies.

In practice we have to limit the number of laboratory measurements. For obvious reasons we can record spectra at only a finite number of angles and for a finite number of incident energies.

The table below lists the radioactive sources used in this test. There is the added complication that these sources do not, for the most part, emit a monoenergetic beam. This causes spectral features to "ride" on top of others complicating the analysis. The MSFC test identification number for these angular response tests is MSFC-BATSE-TPS-19.

Radioactive Sources Used for BATSE Testing			
Isotope	A_Z Symbol	Half-Life	Energies (MeV)*
Americium-241	$^{241}_{95}\text{Am}$	458 yr.	0.060
Barium-133	$^{133}_{56}\text{Ba}$	7.2 yr.	0.032, <u>0.080</u> , 0.276, <u>0.302</u> , <u>0.356</u>
Cadmium-109	$^{109}_{48}\text{Cd}$	453 d.	0.023, 0.088
Cesium-137	$^{137}_{55}\text{Cs}$	30.0 yr.	0.032, 662
Cobalt-60	$^{60}_{27}\text{Co}$	5.62 yr.	<u>1.173</u> , <u>1.332</u> , [2.505]
Cobalt-57	$^{57}_{27}\text{Co}$	270 d.	<u>0.014</u> , <u>0.122</u> , 0.136
Selenium-75	$^{75}_{34}\text{Se}$	-	0.121, 0.265, 0.401
Sodium-22	$^{22}_{11}\text{Na}$	2.62 yr.	<u>0.511</u> , <u>1.275</u> , [1.786]
Thorium-232	$^{232}_{90}\text{Th}$	1.4×10^{10} yr.	2.615, others

* underscore indicates coincident gammas, square brackets indicate sum peaks.

Figure 8 shows the angles chosen for measurement. The regions around 90° and 270° are of particular importance, so they received special attention. At each of these 40 angles a five-minute spectrum was recorded for each source. A background spectrum was then subtracted before storing them on $5\frac{1}{4}$ -inch floppy disks (cf. Appendix A).

2.3. Data Analysis

Figures 9 and 10 show the effects of incidence angle on the spectrum of Cesium-137 as measured by both the Spectroscopy detector (SPD) and the Large-Area detector (LAD). Note the sharper resolution of the spectroscopy detector for the line at 662 keV. Also note the smaller backscatter peak for the very thick spectroscopy detector. Very few photons penetrate through the SPD to be scattered back into the crystal. Also evident in the photopeak of Figure 10 is the slant-path effect mentioned in §2.1. The curve for 45° is not $1/\sqrt{2}$ times that at 0° – as would be expected from a pure $\cos \theta$ decrease.

In order to simplify the study of the dependence on the angle of incidence, we concentrated on three parameters; 1) peak centroid, 2) peak resolution, and 3) peak height. It is obvious that the height and resolution should show a dependence on angle. We expect the peak to decrease in size due to the decreasing intercepted flux. Further, we would expect the resolution ($R = \text{FWHM}/\text{Energy}$) to become degraded (i.e., increase) as the angle gets closer to 90° . These effects are seen. However, it is not obvious that the peak centroid should show an angular dependence. In the first approximation, the peak centroid reflects the energy of the incident gamma ray. Since this energy is *independent* of the angle of incidence, the angle is not expected to have an effect on the position of the centroid. That would be true if the crystal were spatially homogeneous in its response to energy deposition. This apparently is not the case.

2.3.1 Radial Response

In actual fact, the detectors are known to possess what we have termed a “radial response” function. This function represents the fact that energy deposited near the perimeter of the detector has a light collection efficiency that is roughly 15% less than for that deposited near the center. This is not a property of the scintillator, but rather, a property of the light-collection process. Therefore, when the crystal is illuminated, the output is a superposition of gaussians, not centered about the true centroid, but shaded to the left (i.e., lower energies).

The most interesting effects of this radial response occur for angles of incidence near 90° . First, since the photons that are totally converted in the crystal are deposited near the edge, the system responds with a gaussian that is representative of that region near the perimeter. The output therefore indicates a lower energy than that of the incident photon. The result is a photopeak that is lower in energy so its centroid is shifted to the left of where it should be.

In addition to the lowering of the output peak centroid energy, the output does not contain the large variation in centroids which are present when the illumination

is normal (i.e., 0°). The result at $\theta = 90^\circ$ and 270° is a photopeak with a truer gaussian shape.

Both of these effects are seen in Figures 11 and 12. These present the 662 keV peak at angles of 0° and 266° . Note that at 0° the peak is a superposition of many different gaussians representing the contributions from all parts of the crystal. The theoretical gaussian shown in Figure 11 has a centroid at 662 keV. The data at 266° , on the other hand, peak at a lower energy, are less dispersed, and show a purer gaussian shape because they arise from a more localized region of the crystal.

2.3.2 Results

Figure 13 presents the centroid of the photopeak as a function of angle of incidence. Note the sharp drop in position near 90° and 270° for the cesium peak. We have also plotted on this graph partial measurements for lower energy radiation. For the 88 keV and 265 keV lines it was difficult to measure the peak centroids at angles greater than 80° because of blending or lack of penetration through the back of the detector. Still the trends of these two are similar to that of the cesium line.

We have also measured the resolutions (FWHM/Energy) of the cesium peak for all 40 angles. Figure 14 presents the resolution vs. angle. Since the resolution of a peak is a measure of how spread out it is, we find that this parameter decreases (improves) near 90° .

Since these data were all for module 1 (LAD), the table below presents the resolutions for all of the modules (LAD and SPD) for 88 keV and 662 keV lines respectively at 0° .

Instrument Resolution (FWHM %)			
DM #	LAD (88 keV)	SPD (662 keV)	
	@MSFC	@MSFC	@UCSD
1	27.3	7.3	7.7
2	23.6	7.6	7.3
3	24.5	8.1	7.3
4	26.8	7.3	7.4
5	28.5	7.2	6.9
6	24.9	7.1	6.9
7	-	7.5	7.7
8	25.6	7.2	-

Finally, Figure 15 presents the measurements of peak height vs. angle. As mentioned above, the principal features are the cosine-like decrease for angles between 0° and approximately 75°, the presence of deposited energy at 90°, and the attenuated deposition at angles greater than 90°. The ability of the higher-energy gammas to penetrate through the backside of the detector is evident (compare Cs-662 keV vs. Ba-256 keV).

3. Performance Tests: TPS-59

Other tests were performed on the detectors (at normal incidence) to measure such things as detector resolution, energy vs. channel linearity, and to note any inter-detector differences. These tests carry the identification TPS-59.

As an example, Figures 16 and 17 show the differences in the recorded spectra for Large Area detectors 0, 2, and 8. In Figure 16 module 8 shows considerably more dispersion than do the other two modules. This reflects a greater lateral inhomogeneity in that detector's response to the conversion of gamma-ray energy into an electrical output. At lower energies however, (cf. Figure 17) the differences are less. This large radial response is not too disconcerting. The LAD is primarily a low-energy instrument (~100 keV). The performance of DM8 at 1275 keV is well within specifications.

Figures 18 and 19 show that the detectors are very linear in energy per channel. Although this figure presents results only for module 1, the other modules demonstrate the same linearity. For future reference, the equations for energy (in keV) as a function of channel number for modules 1, 2, and 5 for both the LAD and SPD are listed below.

$$\begin{array}{lcl}
 E = 4.9504 \cdot \text{chan.}\# - 28.042 = 4.9504 \cdot (\text{chan.}\# - 5.664) \\
 E = 5.0221 \cdot \text{chan.}\# - 33.045 = 5.0221 \cdot (\text{chan.}\# - 6.580) \\
 E = 5.1374 \cdot \text{chan.}\# - 34.432 = 5.1374 \cdot (\text{chan.}\# - 6.702)
 \end{array} \left. \vphantom{\begin{array}{l} \\ \\ \end{array}} \right\} \text{LAD}$$

$$\begin{array}{lcl}
 E = 3.9931 \cdot \text{chan.}\# - 154.21 = 3.9931 \cdot (\text{chan.}\# - 38.62) \\
 E = 3.9456 \cdot \text{chan.}\# - 154.21 = 3.9456 \cdot (\text{chan.}\# - 37.45) \\
 E = 4.1855 \cdot \text{chan.}\# - 164.52 = 4.1855 \cdot (\text{chan.}\# - 39.31)
 \end{array} \left. \vphantom{\begin{array}{l} \\ \\ \end{array}} \right\} \text{SPD}$$

The coefficients above were calculated from measurements of photopeak centroids by the GETAPEAK program (cf. Appendix B.). These centroids for module 1 are shown in Figures 18 and 19. They are derived from a gaussian fit to the right-hand side of the data peak as demonstrated in Figure 11. It is our intention to repeat these measurements and the linear regression using the centroids of the whole peak. This will certainly result in different values for the coefficients.

4. Conclusion

The measurement of gamma-ray spectra at many different angles and energies is necessary for a correct interpretation of future data from sources of unknown origin. We have discovered that one important facet of the angular response is the presence of the detector's "radial response" first mentioned in §2.3.1. This variation in crystal response as a function of distance from the center must be accurately modelled if future analyses are to be successful.

We have begun to calculate the net response of the crystal under the assumption of different response functions. We have assumed, as first approximations, that the radial response changes from 100% in the center to 85% at the edge and that this variation has a radial dependence that is 1) linear, 2) quadratic, and 3) (co)sinusoidal. The results from these calculations are in a preliminary stage and have not been included in this report. The true radial dependence could very well be a non-analytic function. Tests on the detector to be performed this fall will attempt to actually measure this function. We will then attempt to fit analytical curves to the measurements to include in future calculations.

Bibliography

- 1) Harding, A. K., V. Petrosian, and B. J. Teegarden, 1986, Spectra and Emission Mechanisms, Ch. 2 in *AIP Conference Proceedings*, **141**, *Gamma-Ray Bursts*, ed. by E. P. Liang, AIP, New York.
- 2) Price, W. J., 1964, Nuclear Radiation Detection, Second Edition, McGraw-Hill Book Company, New York.

Appendix A:

A.1 Importing the ND-76 RX-50 disks to the MSDOS format.

The Nuclear Data 76 is the instrument which records the spectra measured by the detector. It is a multichannel analyzer with an analog to digital converter (ADC). The recorded spectra can be saved on $5\frac{1}{4}$ -inch disks. The format of these disks is compatible with the RX-50 standard set by the Digital Equipment Corporation.

Even though most of our data analysis will be performed on the center's main-frame computers, our microcomputers offer several advantages in their ease of use, availability of software, and speed. However, the ND-76 disks are not compatible with the MSDOS standard.

There are two methods that we considered to import the files to the PC's. The first was to try one of the many multi-format programs that allow a PC to read many different formats including CP/M, TRS-80, and Apple disks. Since the Nuclear Data disks are single-sided, quad density, this precludes the use of the 360-kbyte disk drives usually found in the IBM PC. However, the PC/AT compatibles come with 1.2 Mbyte drives. These can read the higher density. Unfortunately, none of the programs that we tested had the RX-50 format as one of its choices. This would have been the most straightforward method. We are still looking into this possibility.

The second method involved reading the ND disks directly in the RX-50 disk drive of the BATSE μ VAX. Because of the DEC compatibility this was an easy task. With a disk inserted in floppy drive #1, log onto the destination directory and issue the following commands;

```
EXCHANGE
COPY $FLOPPY1:*. * /TRANSFER=BLOCK
EXIT.
```

The files are organized in subdirectories on the VAX under the BATSE account name. The subdirectories are named DM0, DM1, ..., for the Detector Modules. Within each subdirectory there is a further set of subdirectories representing the different tests run on that module. So, for example, the spectra taken during the TPS-19 tests of module 1 are found in the directory \$DISK2:[BATSE.DM1.TPS19]. VAX DCL command macros have been written to facilitate movement between these deeply-nested subdirectories. To go to the aforementioned directory, issue the DCL command: CD 1 19.

Getting the files from the VAX onto the PC's was easier than we first thought. Using an Ethernet connection the files were quickly transferred using NFT. For example, to copy all of the TPS-19 files for module 1 the commands are:

```
NFT
COPY BATSE"BATSE password"::$DISK2:[BATSE.DM1.TPS19]*.* *
EXIT
```

Once on the PC, these files in binary format are converted into ASCII files with the CONVERT program written by the author (cf. Appendix B). The spectra were then archived on 1.2 Mbyte floppy disks.

Appendix B.

B.1 Software Development

We have written two programs for use by the BATSE team in their analysis of the test data. The first, called CONVERT, converts a spectrum from the binary format used by Nuclear Data to a straight ASCII format. These converted files can then be read into Lotus 1-2-3, or other programs. Since the MSDOS machines allow 8-character filenames (vs. 6 for the ND), we followed the following algorithm for choosing the output file name. Two examples of spectral files are LA1165.CSL and SPD1VT.NA0. The first is a (TPS-19) measurement by the Large Area detector module 1 of a cesium-137 spectrum at an angle of 165°. Since the 'L' in 'CSL' is redundant, we dropped it to form an 8-character MSDOS filename with an extension of '.prn'. In the second case, from TPS-59, the first three letters were always LAD or SPD for either the large area detector or the spectroscopy detector. We therefore dropped the 'D'. The two output files are then LA1165CS.PRN and SP1VTNA.PRN.

B.2 CONVERT.EXE

The user of CONVERT has three choices depending on the type or number of data file(s) that he is converting.

- 1) The simplest is one input file and therefore one output file. The user is prompted for both filenames. We recommend consistency with the above file-naming convention.
- 2) The second method is used for angular response files. Here there are up to 40 filenames representing the 40 angles of detector orientation. If this option is chosen, the program reads a data file (ORIENT.DAT) containing the 40, or less, angles. It also prompts the user for other pertinent information such as module number, detector, and source.

- 3) The final choice is used for TPS-59 data files. In this case, the root filename is constant and only the extension changes. The user is prompted for the root (e.g., LAD1VT) and the extensions are stored in a data file called 'EXTENS.DAT'.

B.3 GETAPEAK.EXE

The second program, called GETAPEAK, allows the investigator to measure the properties of spectral photopeaks. We have used it to measure centroids, resolutions, and peak heights. It also does an automatic fitting of a gaussian to the peak or portions of a peak. Further information can be gotten by running the program and pressing the F1 (HELP) key.

We plan revisions to both of these programs. The online help should be checked for any changes made after this report is filed.

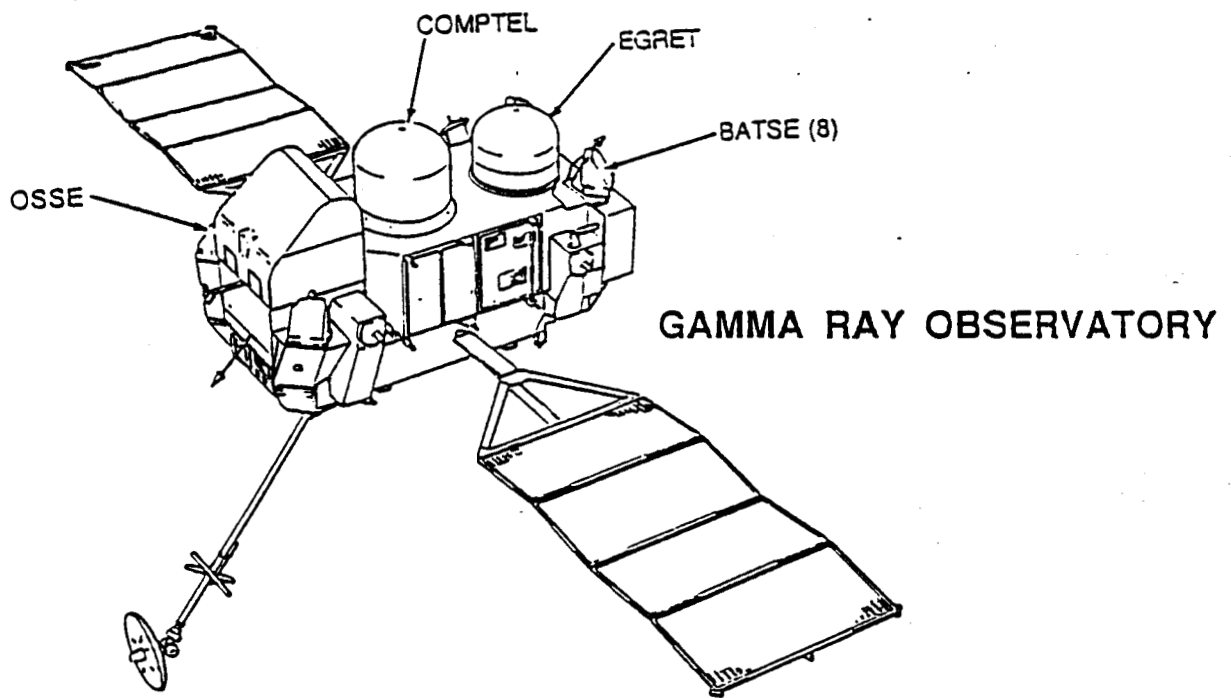


FIGURE 1

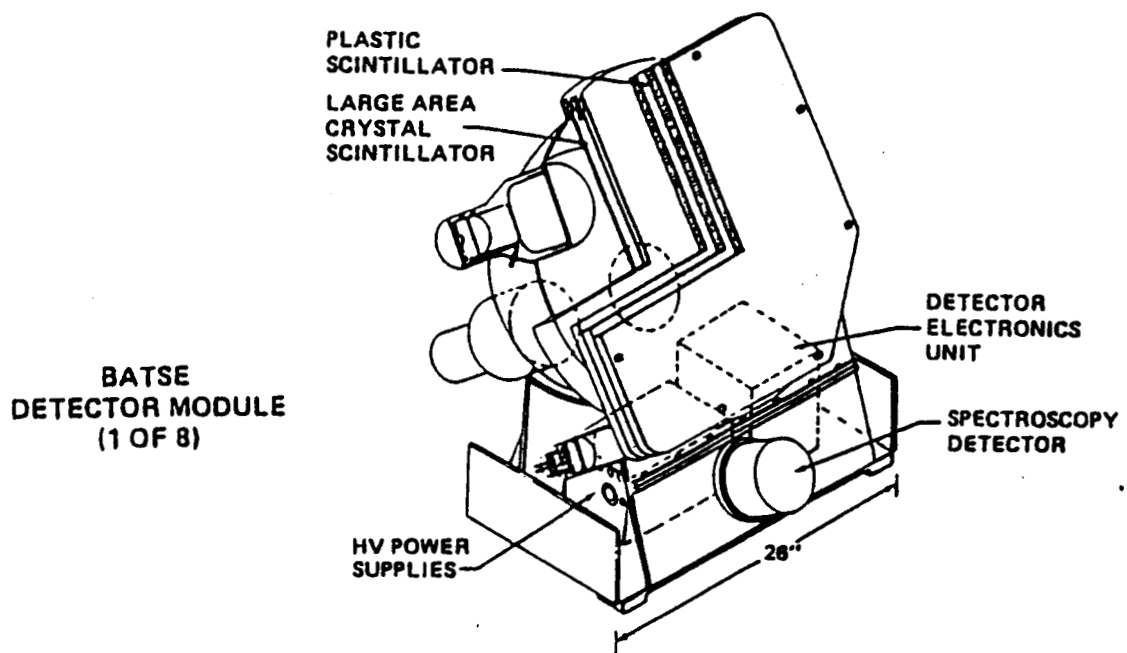


FIGURE 2
XIX-14

ORIGINAL PAGE IS
OF POOR QUALITY

Cesium Spectrum

Module 1 LAD (TPS-19)

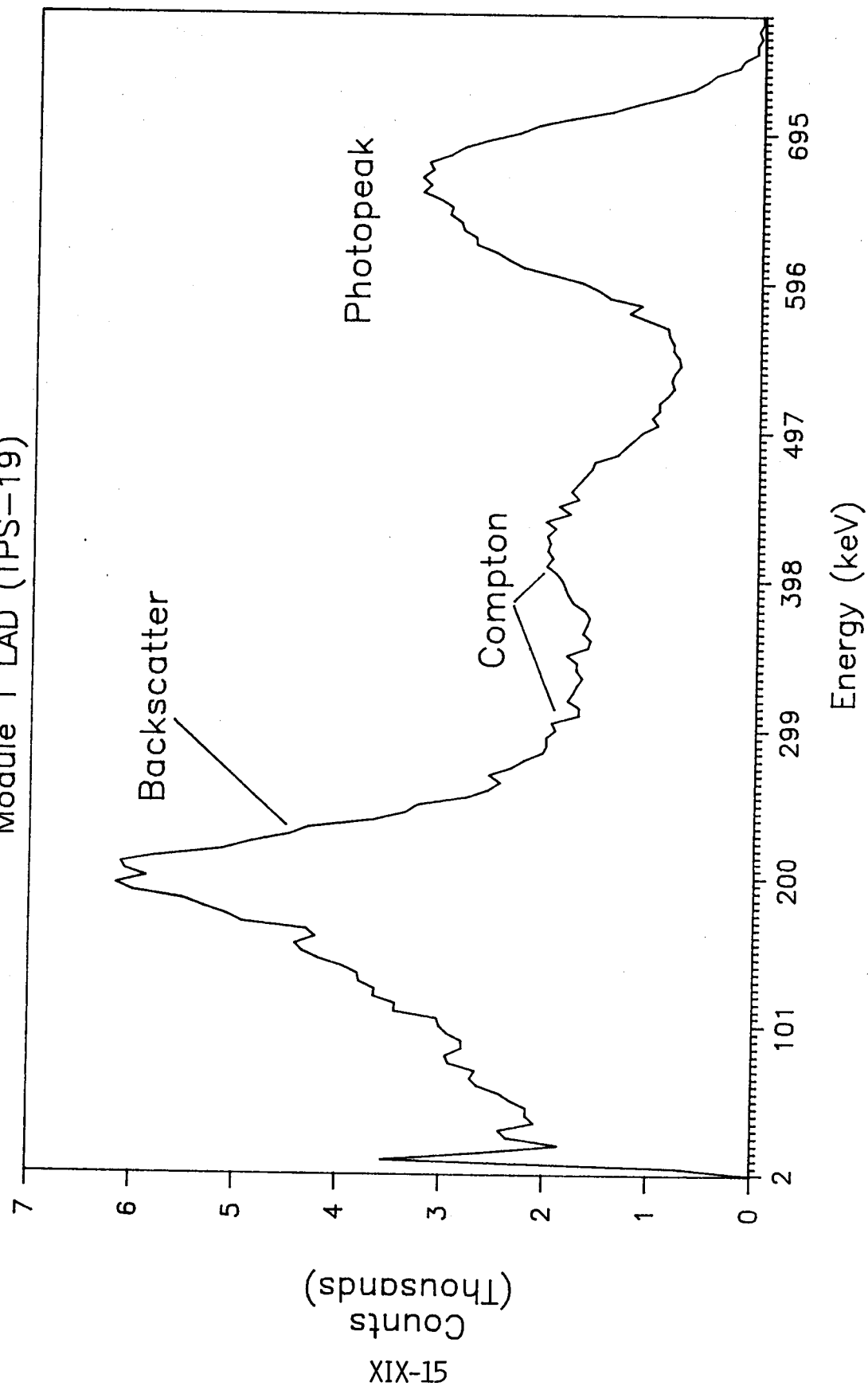


FIGURE 3

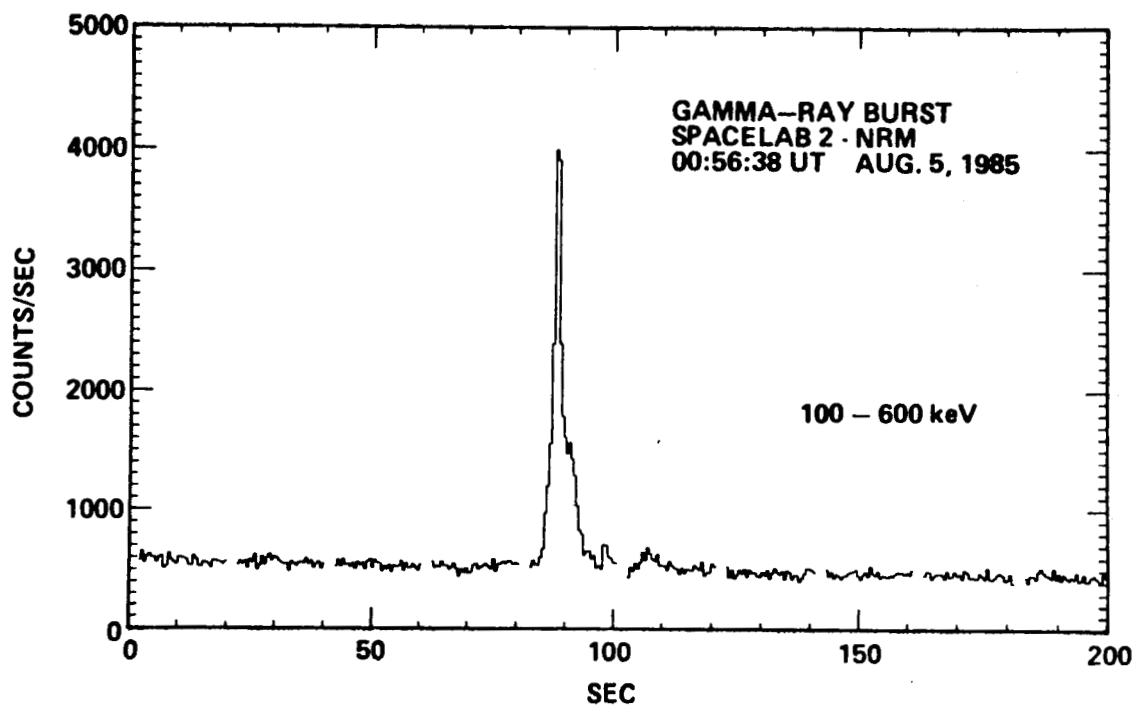


FIGURE 4

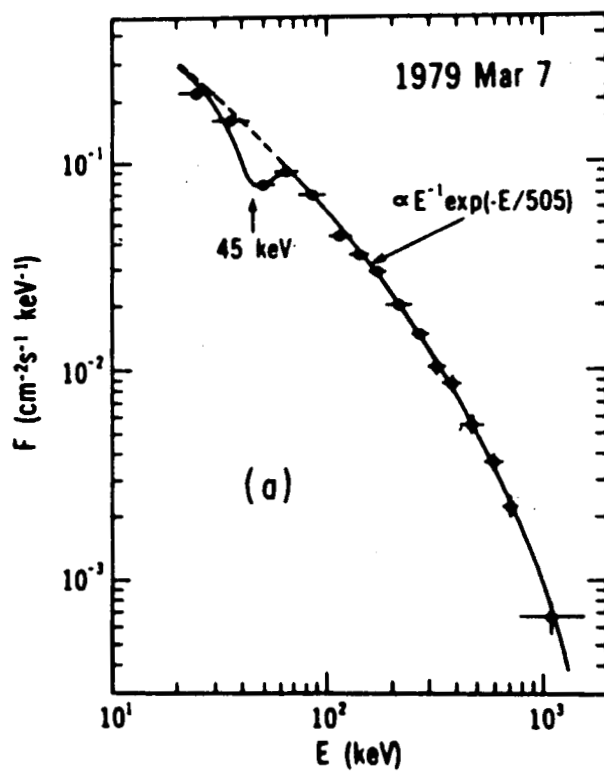


FIGURE 5

The effect of slant path on efficiency

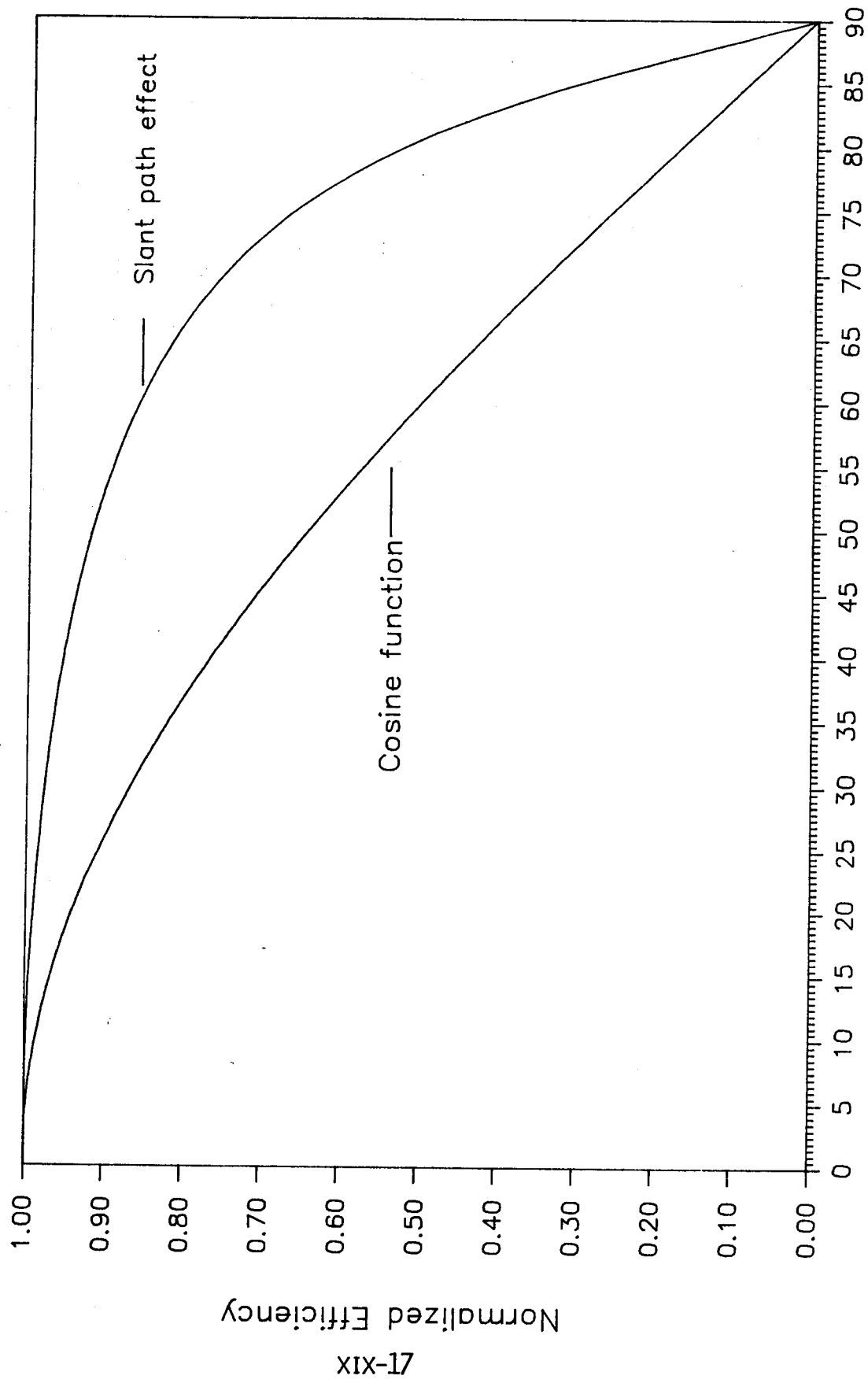


FIGURE 6

Peak height vs. Angle

Cesium 662 keV

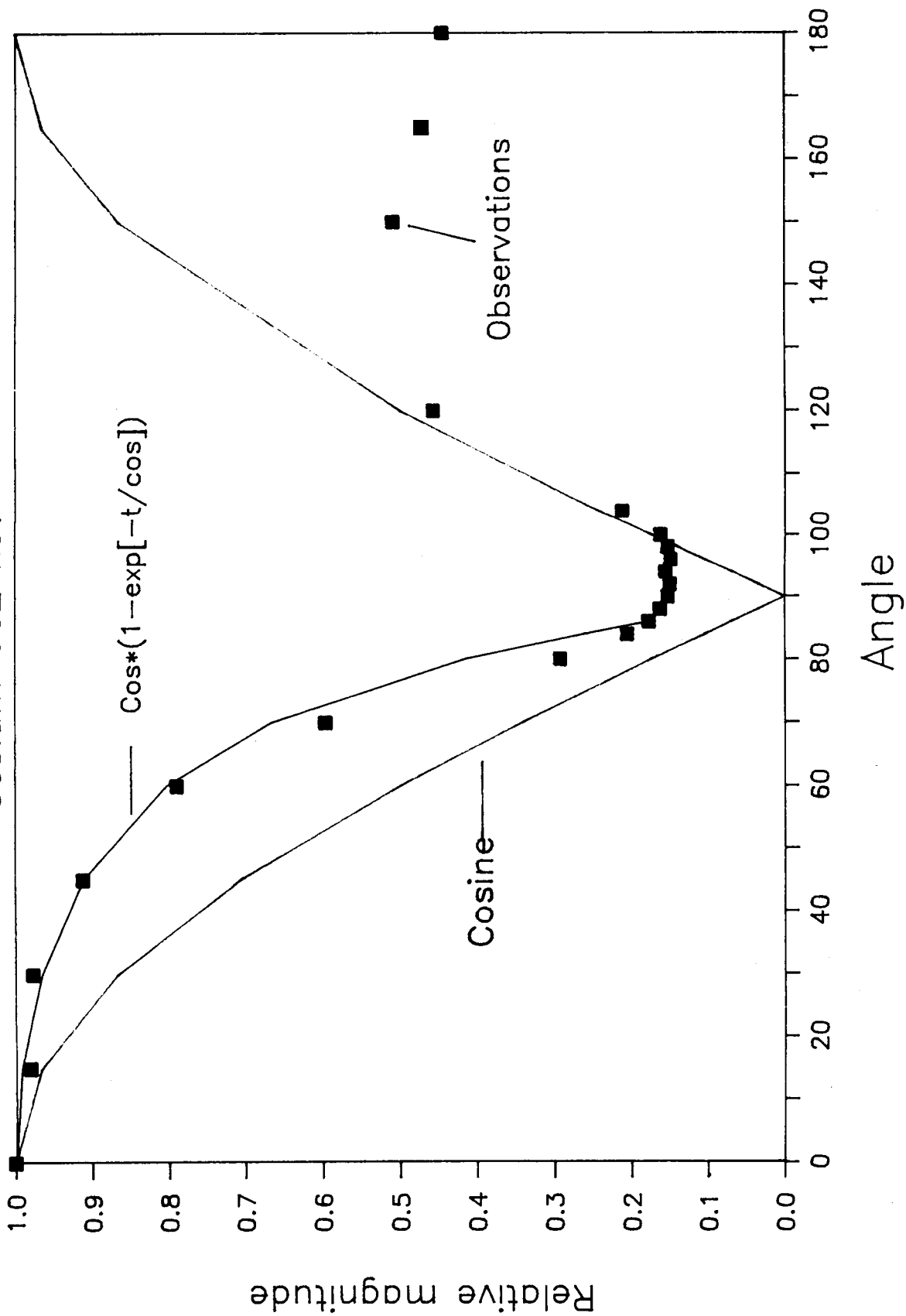


FIGURE 7

Angles of Incidence
for TPS-19

0 degrees

270

90

180

FIGURE 8

XIX-19

Cesium Spectra

Spectroscopy Detector (#1)

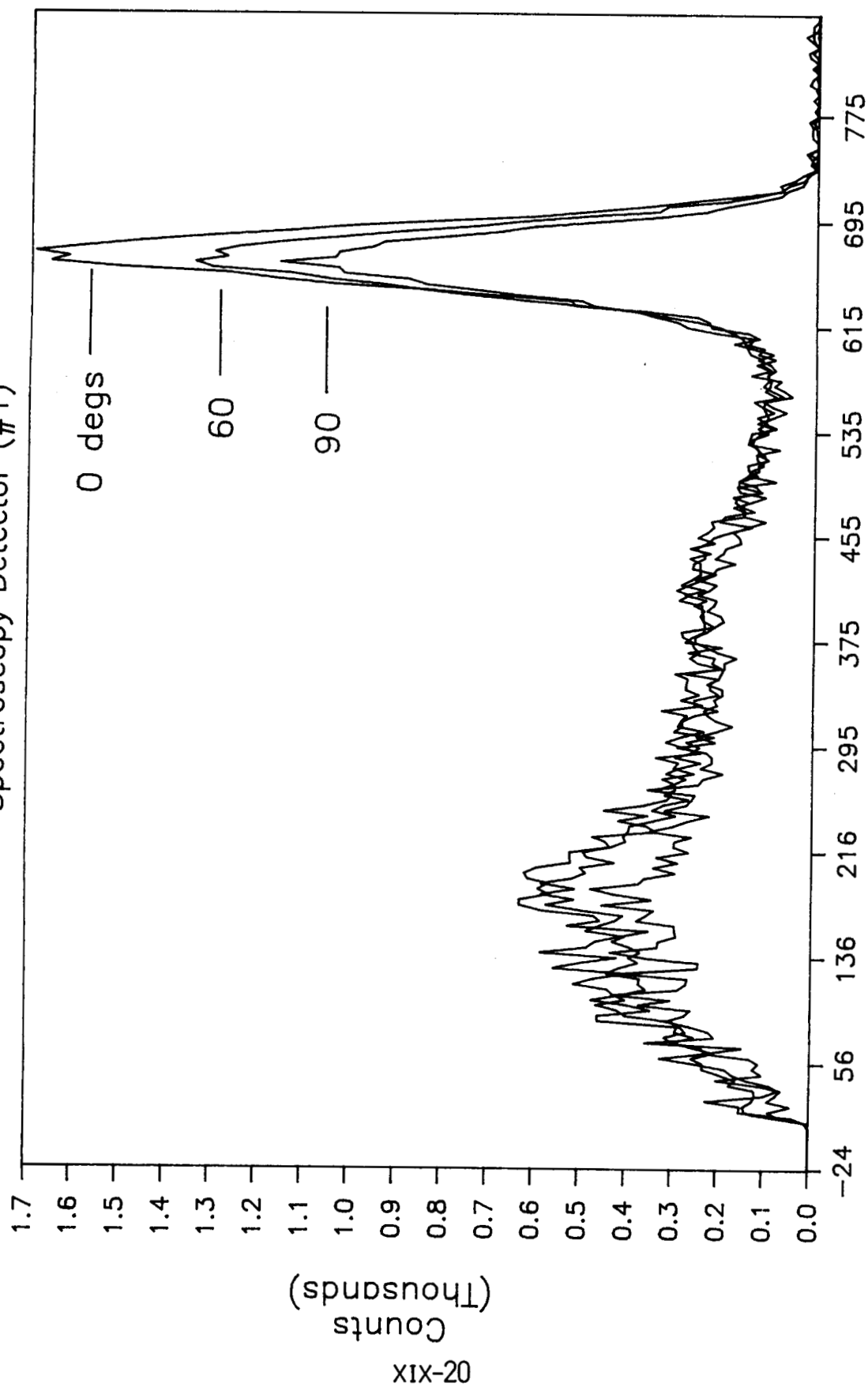
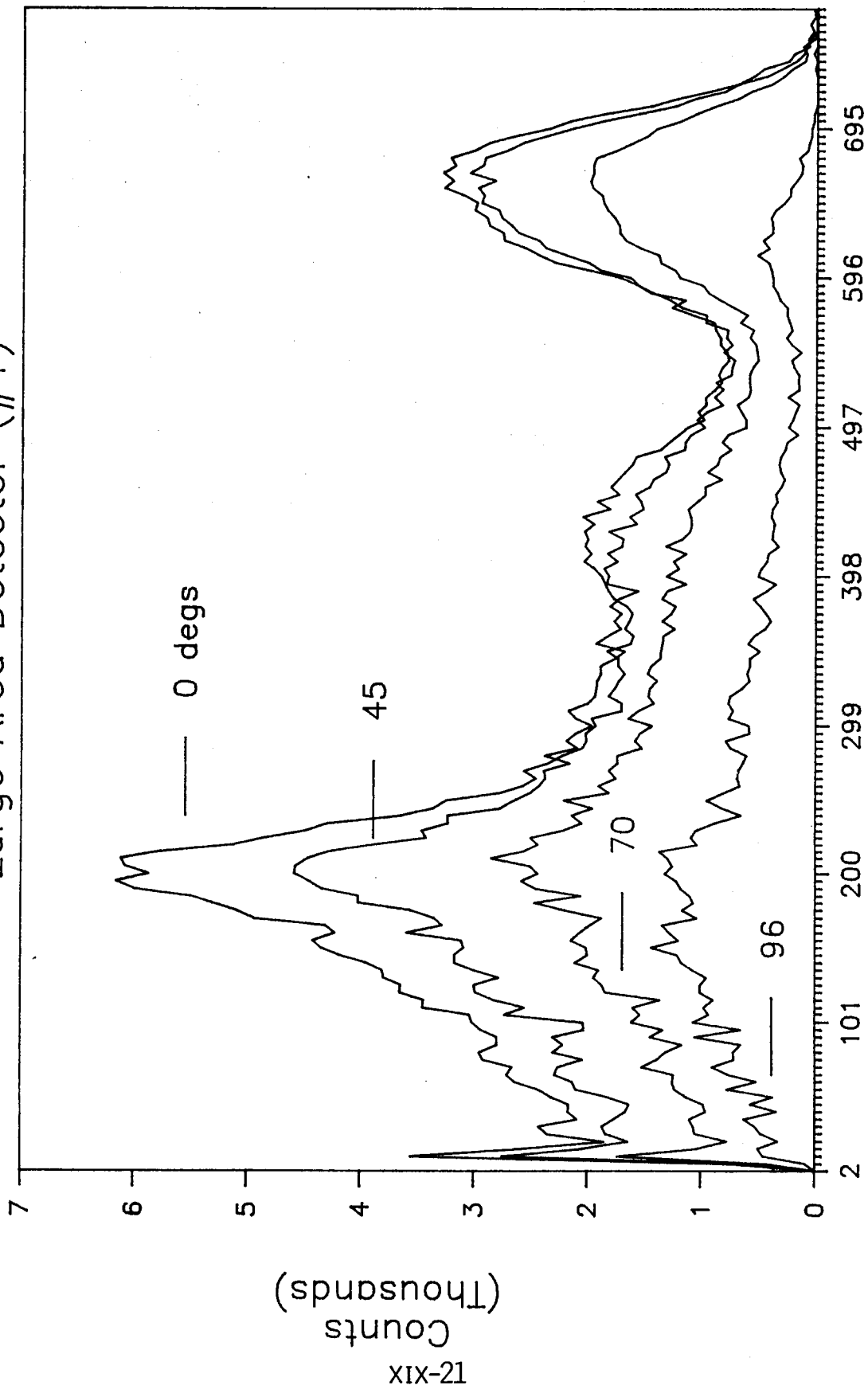


FIGURE 9

Cesium Spectra

Large Area Detector (#1)

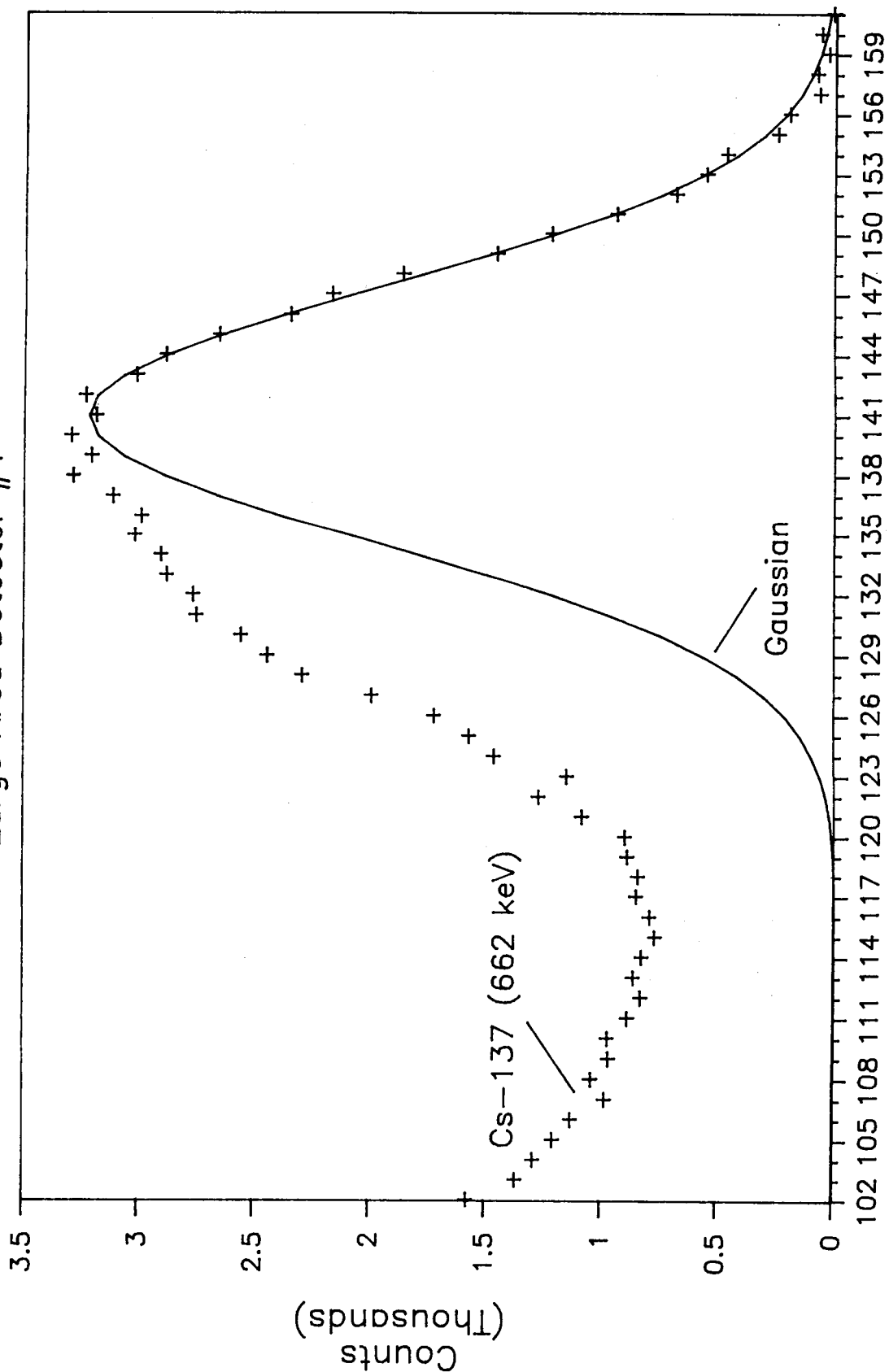


Energy (keV)

FIGURE 10

Gauss Fit

Large Area Detector #1

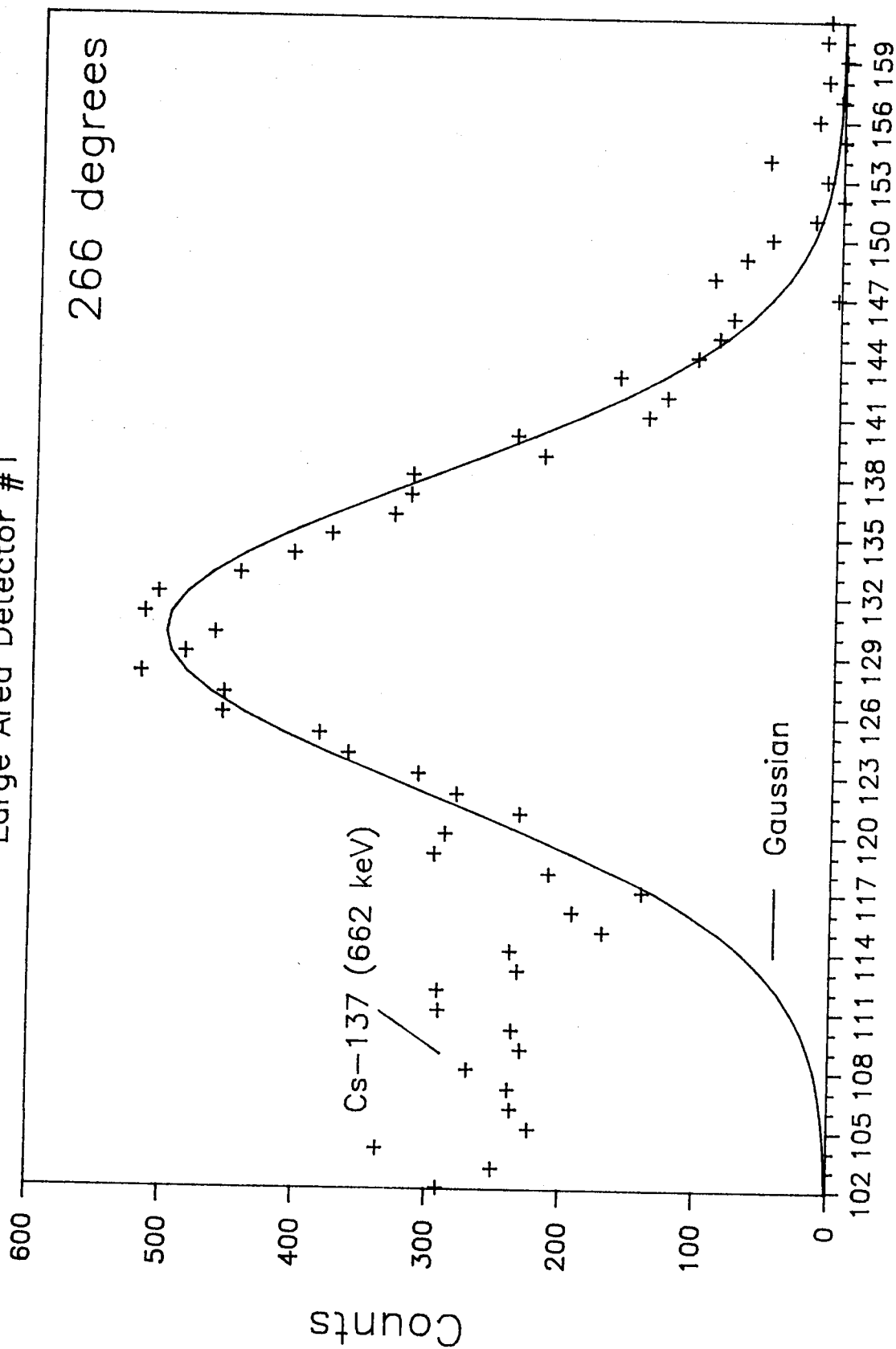


Channel

FIGURE 11

Gauss Fit

Large Area Detector #1



Channel

FIGURE 12

Centroid Shift vs. Angle

Module 1 LAD (TPS-19)

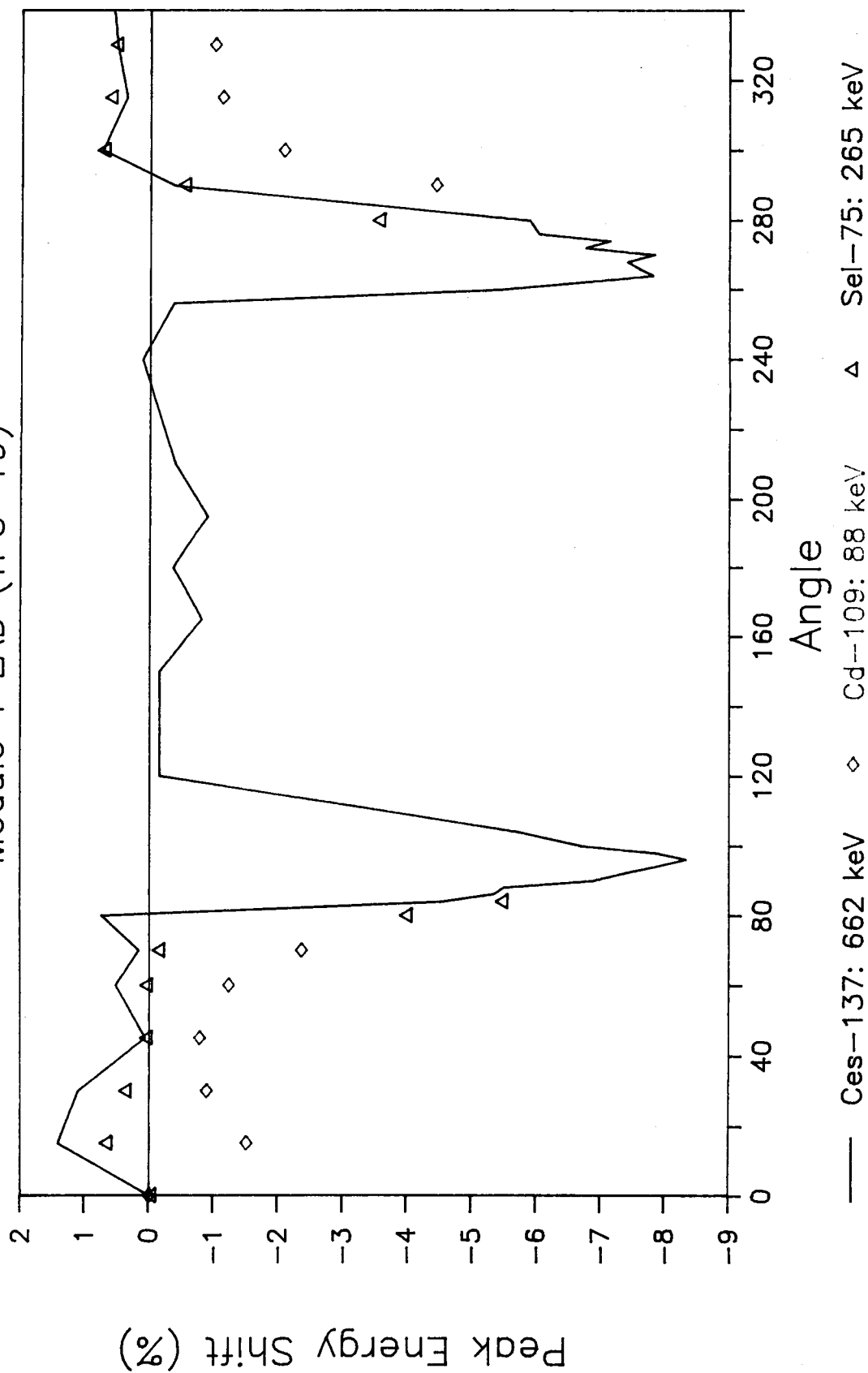


FIGURE 13

Resolution vs. Angle

Module 1 LAD (TPS-19)

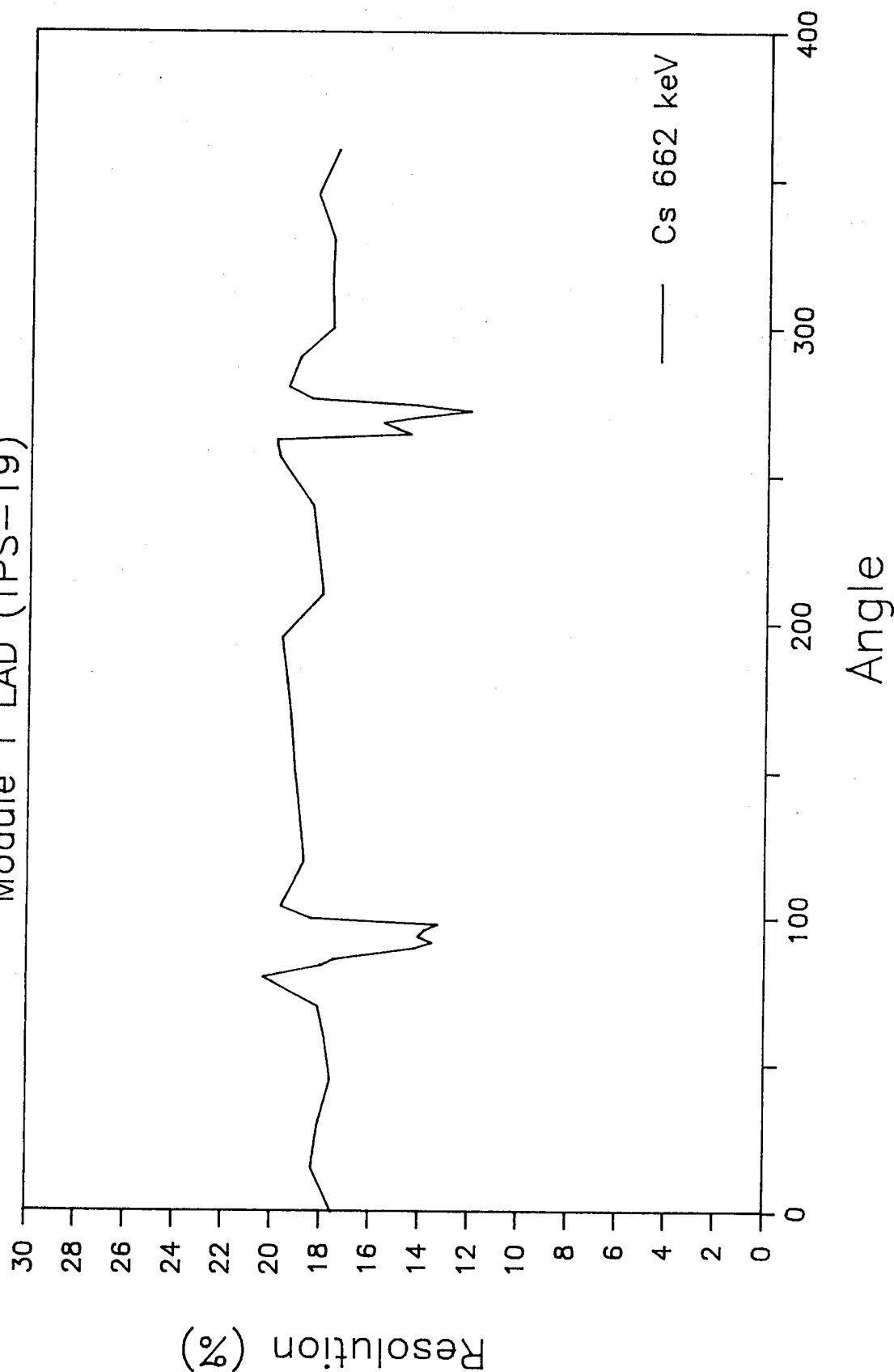


FIGURE 14

Peak Height vs. Angle Module 1 LAD (TPS-19)

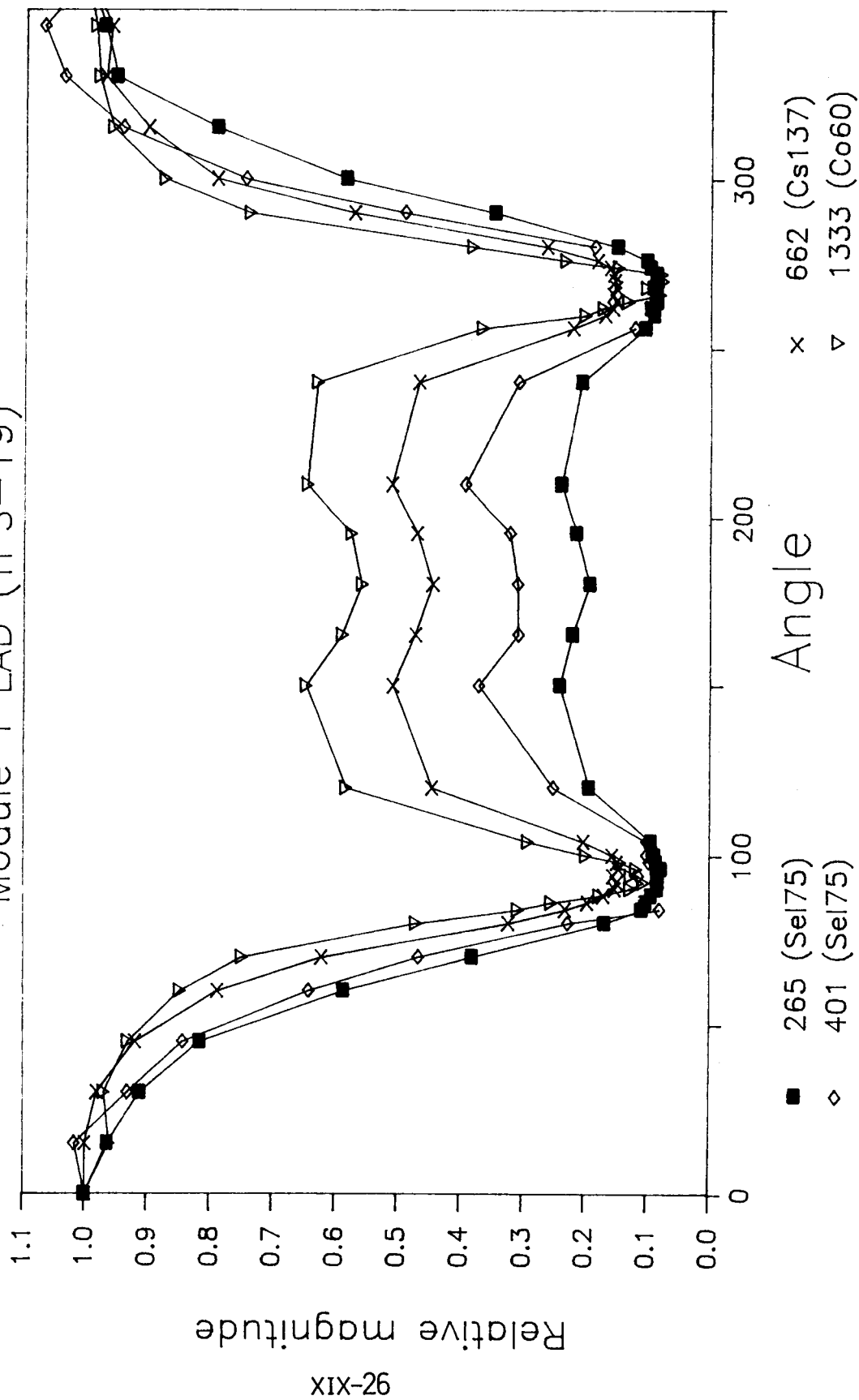
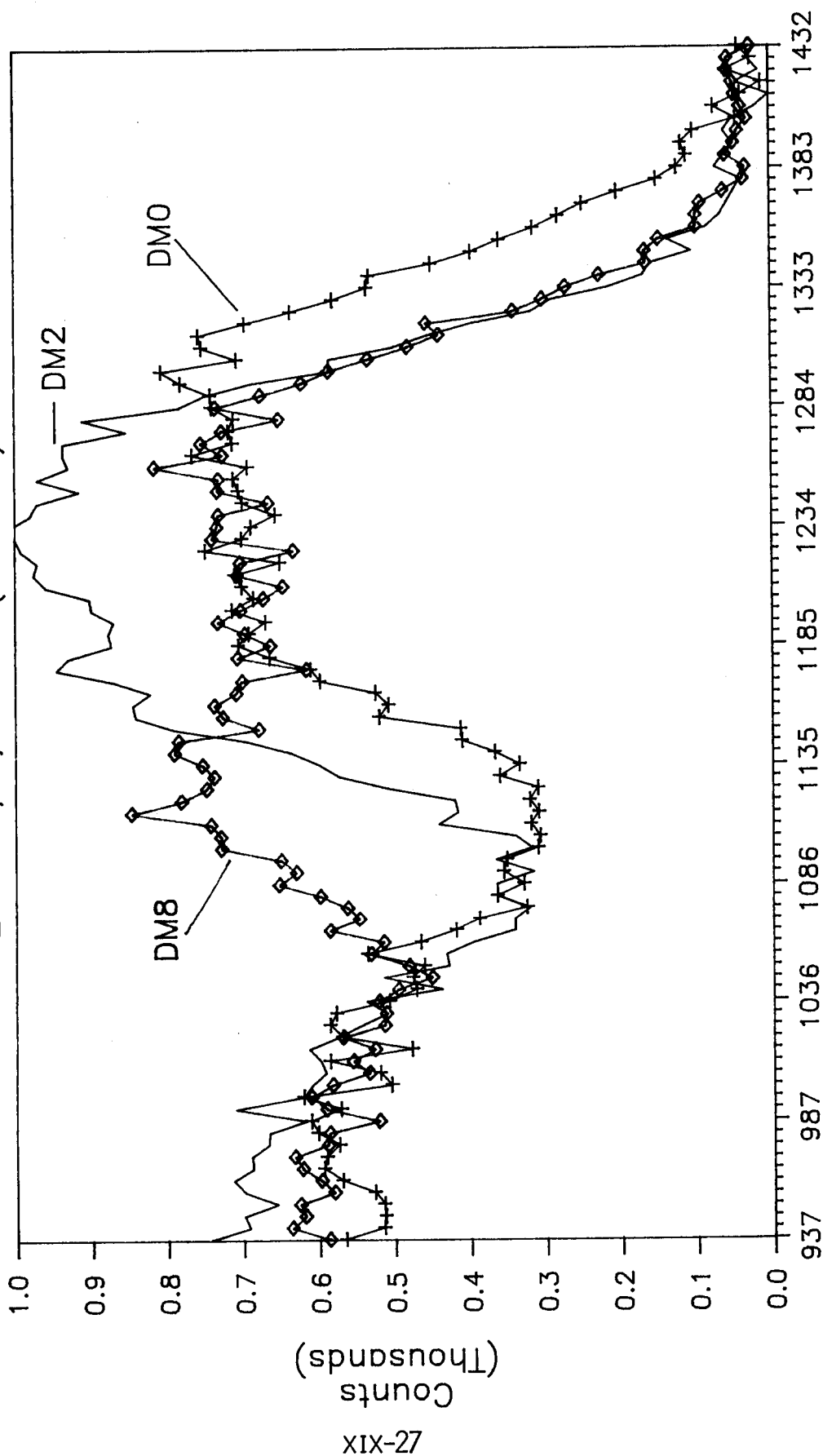


FIGURE 15

Sodium-22 Spectra

LAD's 0, 2, and 8 (TPS-59)

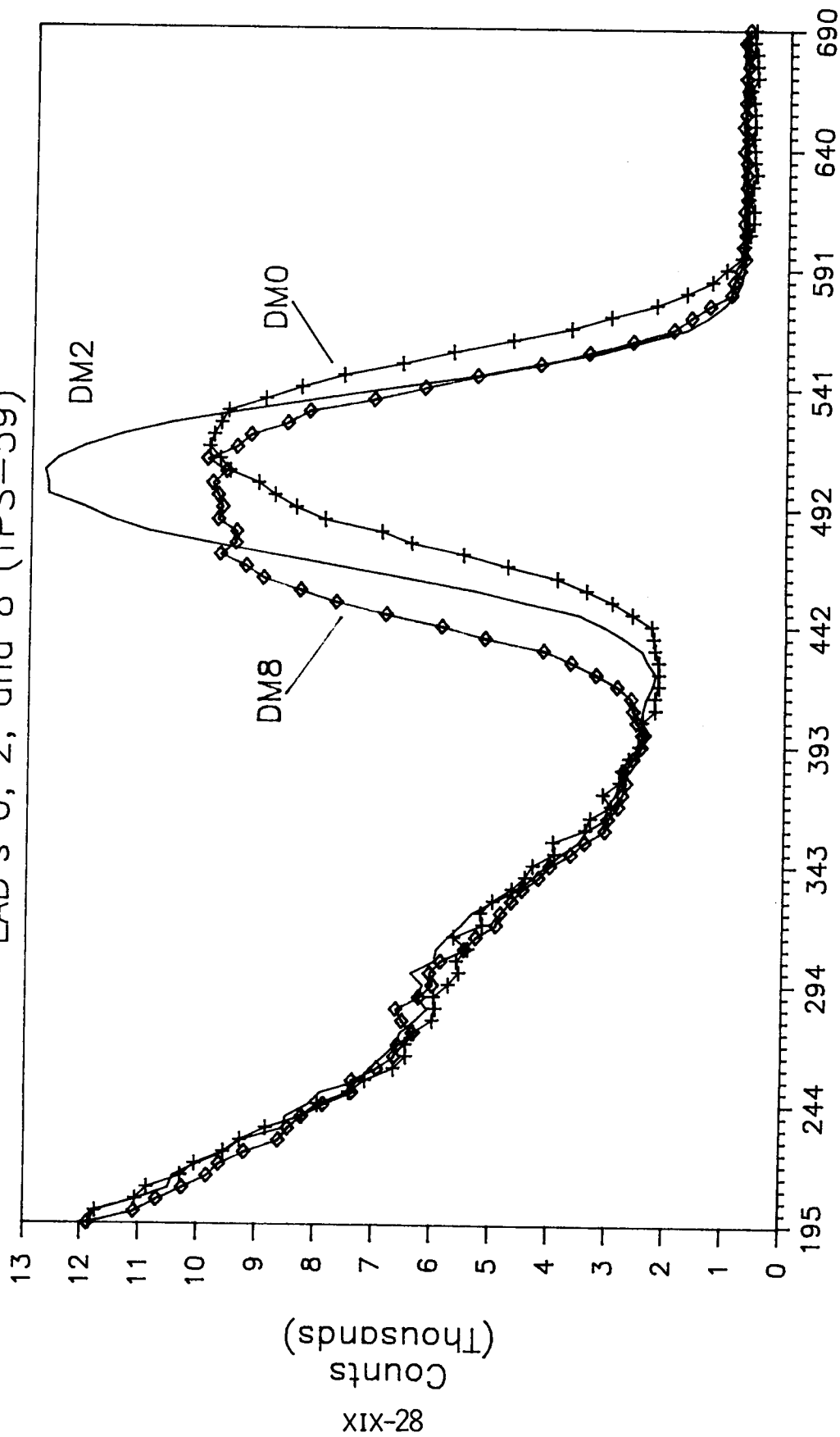


Energy (keV)

FIGURE 16

Sodium-22 Spectra

LAD's 0, 2, and 8 (TPS-59)



Energy (keV)

FIGURE 17

Energy vs. Channel

Module 1 LAD (TPS-59)

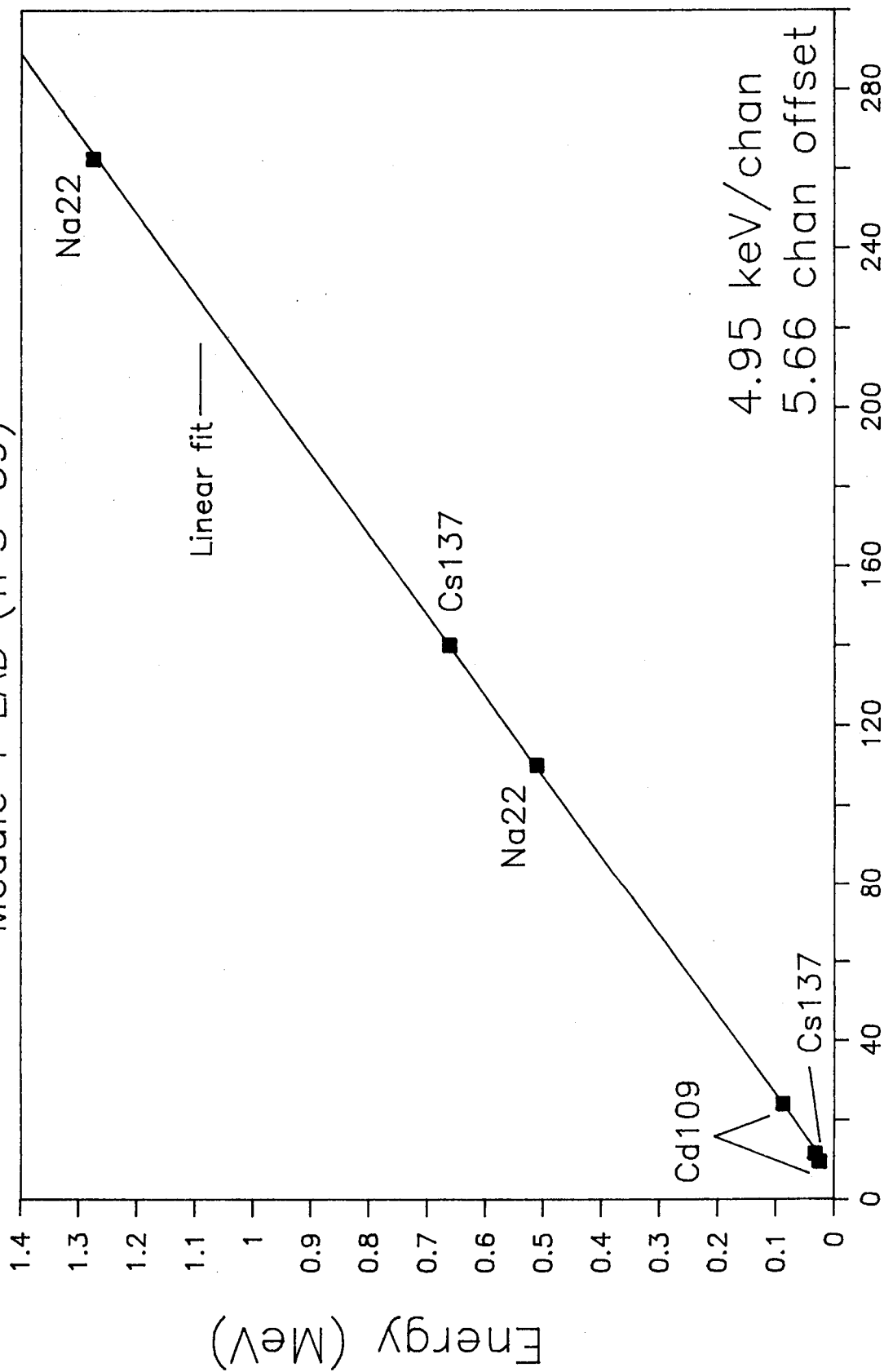
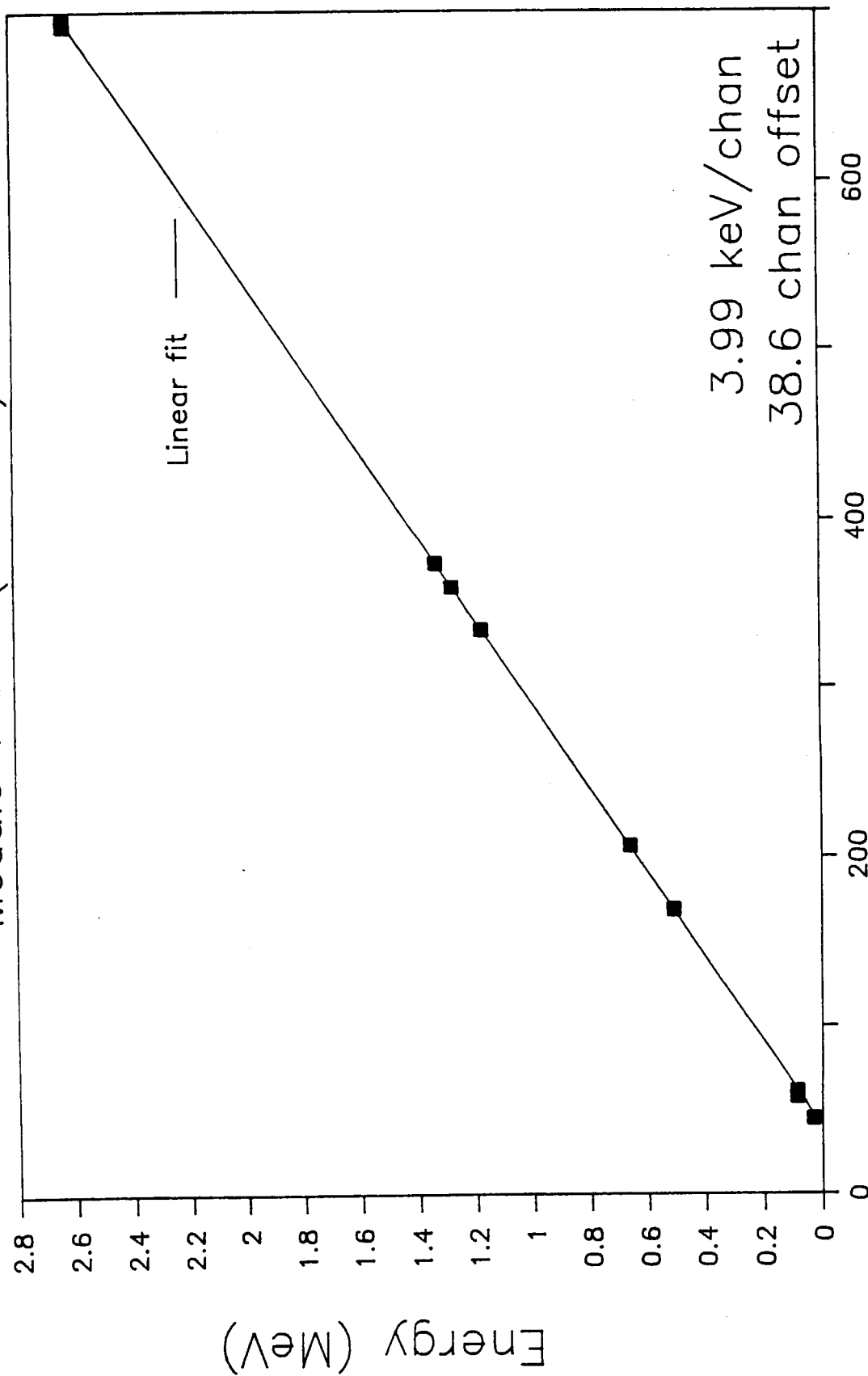


FIGURE 18

Energy vs. Channel

Module 1 SPD (TPS-59)



Channel

FIGURE 19

N89-21746

1988

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

THE SIMULATION OF THE ALTERNATE TURBOPUMP DEVELOPMENT
HIGH PRESSURE OXYGEN AND FUEL TURBOPUMPS FOR THE
SPACE SHUTTLE MAIN ENGINE USING THE SHABERTH COMPUTER PROGRAM

Prepared by: Gary H. McDonald, Ph.D, P.E.

Academic Rank: Assistant Professor

University and Department: The University of Tennessee
at Chattanooga
Mechanical Engineering --
Mechanics

NASA/MSFC:

Laboratory: Propulsion
Division: Component Development
Branch: Turbomachinery and
Combustion Devices

NASA Colleague: Henry P. Stinson

Date: August 19, 1988

Contract No.: NGT 01-002-099
The University of Alabama

THE SIMULATION OF THE ALTERNATE TURBOPUMP
DEVELOPMENT HIGH PRESSURE OXYGEN AND FUEL
TURBOPUMPS FOR THE SPACE SHUTTLE MAIN ENGINE
USING THE SHABERTH COMPUTER PROGRAM

by

Gary H. McDonald
Assistant Professor of Engineering
in Mechanical--Mechanics
The University of Tennessee at
Chattanooga
Chattanooga, Tennessee

ABSTRACT

The space shuttle main engine (SSME) is basically comprised of a combustion chamber and nozzle, high and low pressure oxygen turbopumps and high and low pressure fuel turbopumps. In the current configuration, the high pressure fuel (HPFTP) and high pressure oxygen turbopumps (HPOTP) have experienced a history of ball bearing wear. The wear problem can be attributed to numerous factors including the hydrodynamic axial and radial loads caused by the flow of liquid oxygen and liquid hydrogen through the turbopump's impellers and turbine. Also, friction effects between the rolling elements, races, and cage can create thermally induced bearing geometry changes. If the frictional heat generation becomes greater than the available coolant capacity and if the hydrodynamic loads become excessive, then, then thermal effects and loading can contribute to eventual bearing failure. To alleviate some of the current configuration problems, Pratt and Whitney has proposed the alternate turbopump development (ATD). However, the ATD HPOTP and HPFTP are constrained to operate interchangeably with the current turbopumps, thus, the operation conditions must be similar. The ATD configuration features a major change in bearings used to support the integrated shaft, impeller and turbine system. A single ball and single roller bearing will replace the pump-end and turbine end duplex ball bearings.

In this study, the SHABERTH (Shaft-Bearing-Thermal) computer code was used to model the ATD HPOTP and ATD HPFTP configurations. A two-bearing model was used to simulate the HPOTP and HPFTP bearings and shaft geometry. From SHABERTH, a comparison of bearing reaction loads, frictional heat generation rates, and Hertz contact stresses will be attempted with Pratt and Whitney's analysis at the 109% and 65% power levels.

ACKNOWLEDGEMENTS

Several individuals have provided important contributions in this author's NASA/ASEE fellowship project. I would first thank Dr. L. Michael Freeman, director of the MSFC NASA/ASEE fellowship program and Ms. Ernestine Cothran, director of MSFC university relations, for their diligent efforts in the organization of seminars, tours, and the general administration of this worthwhile program. I thank them for the opportunity to again participate in this program for a second summer.

I would especially like to thank my colleague Mr. Henry P. Stinson who was valuable in formulating this year's project and providing the necessary technical assistance. A special appreciation is extended to Mr. Dale Blount of NASA/MSFC and Mr. John Price and Mr. Yul Eren of Pratt and Whitney who provided the technical information to model the alternate turbopump development design.

Also, a grateful appreciation is extended to Mr. Joe Cody and Mr. Dave Marty of SRS/System Division who provided information concerning the ATD ball bearing for the program's input. Dave loaded both the SHABERTH code and a CLIST execution program on the IBM EADS for the author. Also, an example input file was loaded for the latest version of SHABERTH code which was utilized in this project. Also, Dave supplied helpful information on the use of IBM-Engineering Analysis and Data System (EADS). A special appreciation must be expressed to Ms. Julie Bomar, UTC-School of Engineering, who diligently typed this report.

I finally express my gratitude to Mr. Loren Gross and all the members of the Turbomachinery and Combustion Devices Branch for their technical guidance, assistance, and encouragement that has made my two summers at MSFC an enjoyable and productive experience.

ORIGINAL PAGE IS
OF POOR QUALITY

INTRODUCTION

The space shuttle main engine (SSME) is basically comprised of a combustion chamber, nozzle, high and low pressure fuel and oxygen turbopumps. In the current configuration, the high pressure fuel and oxygen turbopumps (HPFTP and HPOTP) have had a history of ball bearing wear. This wear has contributed to numerous factors including hydrodynamic axial and radial loads caused by the flow of liquid hydrogen (LH2) and oxygen (LOX) through the turbopump's impellers and turbines. Also, wear caused by friction effects creates thermally induced bearing geometry changes. Pratt and Whitney Aircraft [8] has been contracted to develop an alternate design for the high pressure oxygen and fuel turbopumps. This alternate turbopump development (ATD) will hopefully alleviate the bearing wear problems of the current design. However, the ATD turbopumps are constrained to operate interchangeably with current turbopump configurations; thus, the ATD must have the same operating conditions to generate the power levels of the current SSME turbopumps. Since the ATD is in the development phase of design, the first step in proving the feasibility of the proposed changes is analytical. Once an analytical basis is established, then the construction and instrumentation of the bearing testers is completed to provide the necessary test data for rating the bearing's performance. The final phase will be the construction and testing of high pressure fuel and oxygen turbopumps. The testing of the turbopumps in an engine system will generate the necessary database to confirm the ATD design for flight standards.

In the analytical phase, one tool in the prediction of bearing behavior is the use of the SHABERTH (Shaft-Bearing-Thermal) computer program. This program, originally developed by SKF Industries, has been greatly modified by SRS Technologies/System Division of Huntsville, Alabama, to model HPFTP and HPOTP for NASA Marshall Space Flight Center (MSFC). SHABERTH input data necessary for the program's execution is a collection of information concerning the bearing's and shaft's geometry and material properties, the location and magnitude of the loads applied to the shaft and the initial thermal environment of the bearings. The collection of input data was performed in this study for both the ATD high pressure fuel and oxygen turbopump configurations. From this given information, SHABERTH will calculate frictional heat generation, bearing Hertzian contact stresses, bearing clearance changes and bearing reaction loads. These are only a sample of SHABERTH's output results, but will be the chosen results to indicate the bearing's performance in this study.

For sample input and output listings, references [4,5] are useful. These references are the user manuals for the SHABERTH program.

ATD Design

The ATD high pressure oxygen turbopump (HPOTP) configuration is shown schematically in Figure 1. The turbopump consists of the same major components as the current configuration. These components are a preburner impeller, main impeller, and turbine disk with blades which are integrated into a system by a common hollow shaft. The main differences in the ATD HPOTP design are the choice of bearings to support the shaft. Currently, a duplex pair of 45 mm bore ball bearings support the pump end, whereas a duplex pair of 57 mm bore ball bearings support the turbine end of the HPOTP. The 45 mm bore pump end ball bearings have been replaced with a single 60 mm bore ball bearing. On the turbine end, the 57 mm bore ball bearings have been replaced by a single 73 mm bore cylindrical roller bearing.

Also, shown in Figure 1 on the turbine end, a small thrust ball bearing exists. Its purpose is to react any transient unbalance axial load toward the turbine. This thrust ball bearing is to only carry transient axial load since cylindrical roller bearings are, by design, radial load carriers. Unfortunately, this thrust bearing is not included in this simulation. This exclusion is due to insufficient information concerning the thrust bearing's geometry at this time. These axial loads are only significant in the start-up or shut-down of the engines. Since a 109% full power level has been chosen to be simulated in this study, these axial loads will not be considered. So, the exclusion of the thrust bearing should not significantly affect the results of this simulation.

Another difference in the ATD HPOTP is the location of the interpropellant seal package. The labyrinth seals are now between the roller bearing and the main impeller. This allows the roller and thrust ball bearing to be operated in a liquid hydrogen (LH2) environment. Whereas, the HPOTP ball bearing will operate in a liquid oxygen (LOX) in environment.

The ATD high pressure fuel turbopump (HPFTP) configuration is shown schematically in Figure 2. Again, this turbopump consists of the same major components as the current configuration. The HPFTP consists of first, second, and third stage impellers and a turbine blades and disk that are integrated onto a hollow shaft. The ATD HPFTP replaces the 45 mm bore duplex pair pump end ball bearings with a single 63 mm ball bearing and the 45 mm bore duplex pair turbine end ball bearings with a single 73mm bore cylindrical

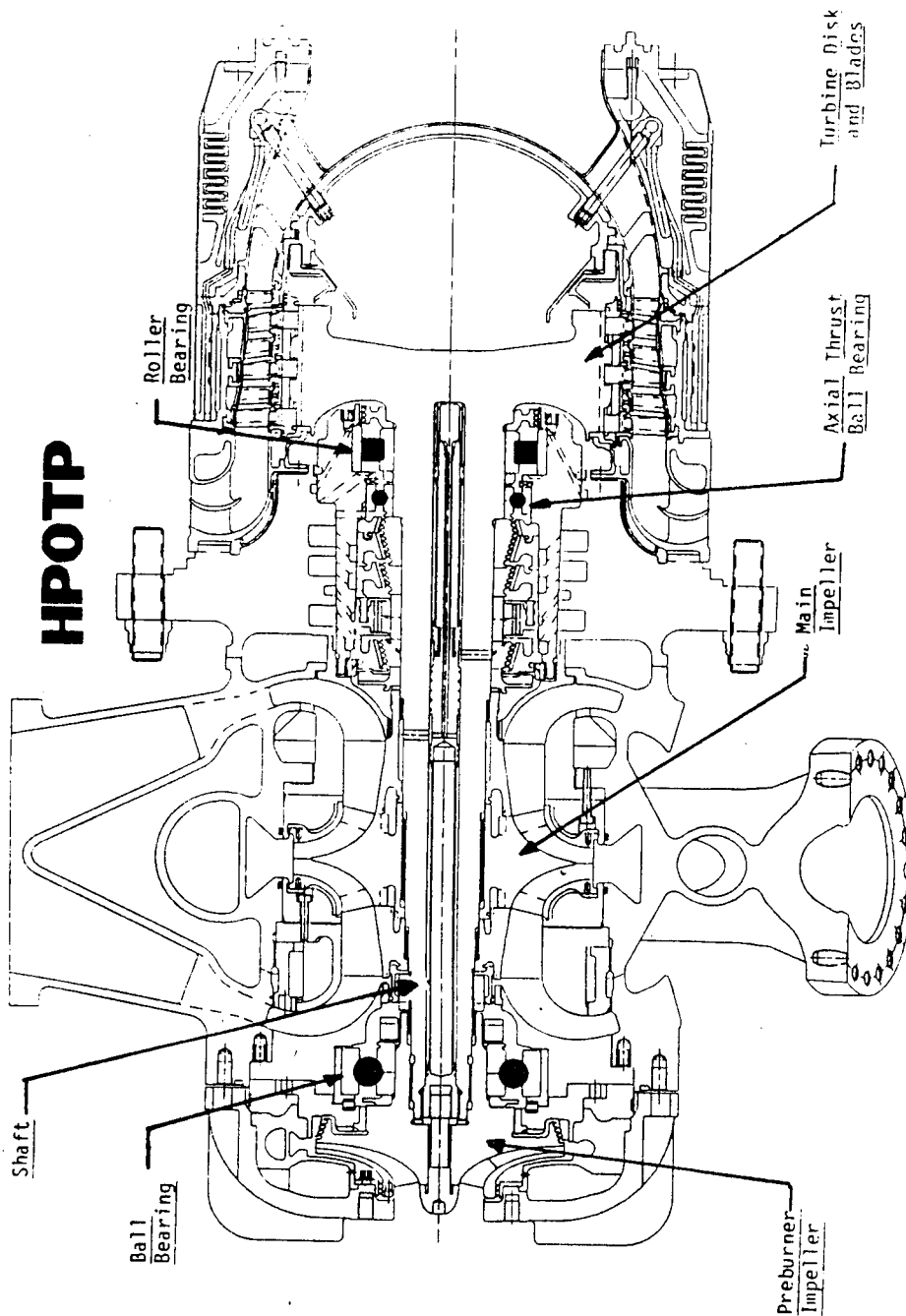


Figure 1: Schematic of the High Pressure Oxygen Turbopump (HPOTP)

HPFTP

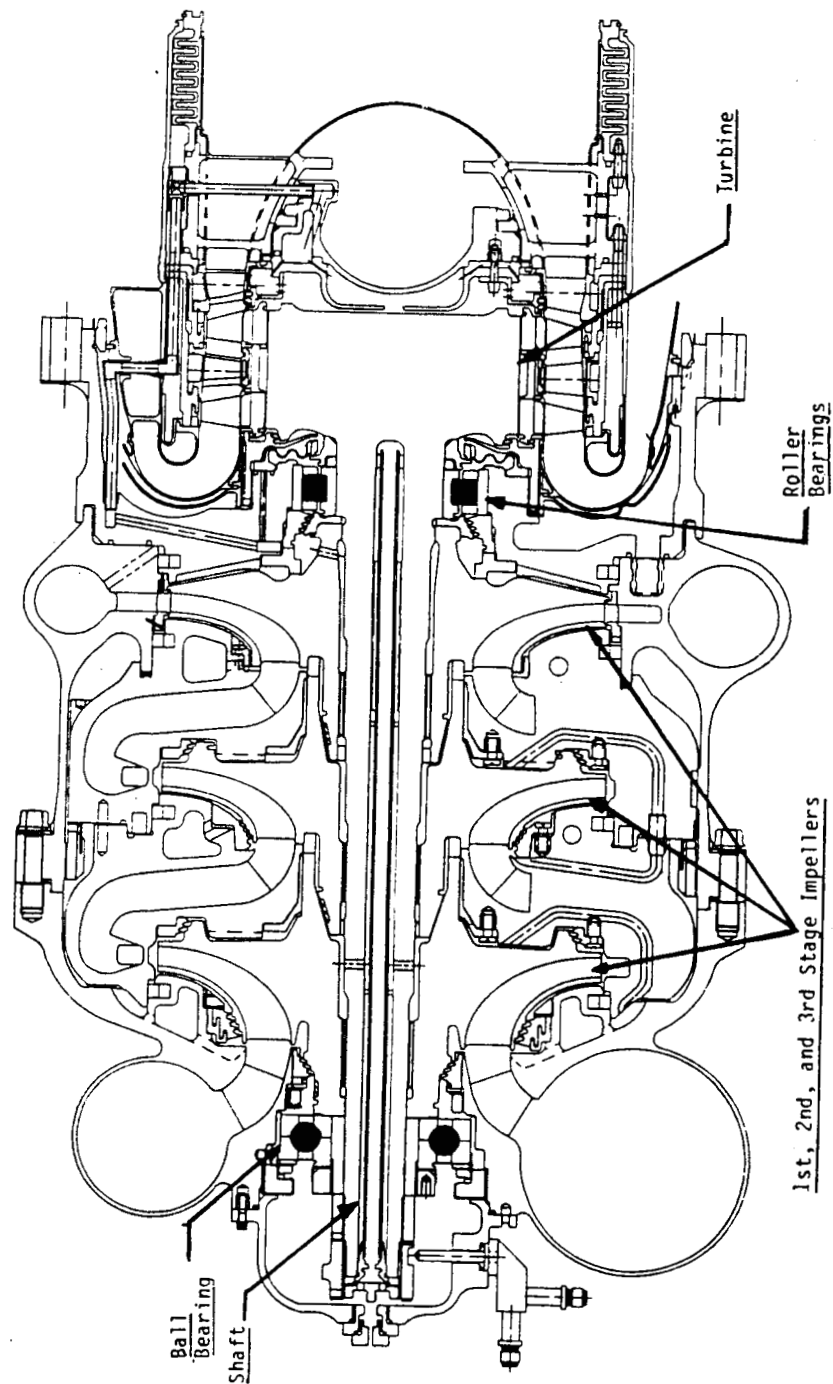


Figure 2: Schematic of High Pressure Fuel Turbopump (HPFTP)

roller bearing. In the HPFTP, no thrust ball bearing is necessary since there exists a balance piston and rub-stops to control any axial movement. Thus, the roller bearing should be reacting to only radial loads. Both the ball and roller bearings will be operating in the liquid hydrogen (LH2) fuel environment.

OBJECTIVES

The purpose of this study is to model the ATD HPOTP and HPFTP configurations using the SHABERTH computer program. Using the results of these simulations, a comparison will be attempted to Pratt and Whitney's predicted values for bearing reaction loads, frictional heat generation and Hertzian contact stresses. Pratt and Whitney performed one bearing models for both the HPFTP and HPOTP ball and roller bearings using SHABERTH. From SHABERTH, they predicted the frictional heat generation, while from the A.B. Jones bearing program, they obtained values for Hertzian contact stress. Therefore, the objectives of this project were

1. To collect radial hydrodynamic load information representing the effects of the major components of the HPOTP and HPFTP for 109% power level and the 65% power level.
2. To create input data files representing the ball and roller bearing geometry and material properties, the shaft size, initial temperatures at 13 locations on the bearing due to coolant flow, coefficients of friction, axial preload for the bearings and the magnitude and location of the hydrodynamic radial loads along the shaft.
3. To attempt a comparison of reaction bearing loads, frictional heat generation and Hertzian contact stress of the two-bearing HPOTP and HPFTP SHABERTH models to Pratt and Whitney or SRS Technologies results.

SHABERTH Computer Model

SHABERTH performs a thermo-mechanical simulation of a load support system consisting of a flexible shaft supported by up to five rolling element bearings. The shaft can be hollow or solid and of arbitrary geometry. Any combination of ball, cylindrical, or tapered roller bearings can be used to support the shaft. The applied loading can consist of point or distributed moments, point or distributed forces and shaft misalignments. Bearing operating clearance is determined as a function of shaft and housing fits. A lumped

mass thermal model allows calculation of steady state or time transient system temperatures considering free and forced convection, conduction, radiation and mass transport heat transfer. The purpose of this program is to provide a tool in which the shaft bearing system performance characteristics can be determined as functions of system temperatures. These system temperatures may be either steady state or transient. The bearing theory used in SHABERTH is found in a reference by Harris [1]. Other bearing theory sources are found in references [2,3]. For a complete discussion concerning SHABERTH's structure, references [4,5] are program's first and second generation user manuals. These manuals provide the user with examples of input format tables and output listings for various cases.

The ball bearing information for both the ATD HPOTP and HPFTP was supplied by SRS Technologies/System Division of Huntsville, Alabama. The ATD HPOTP and HPFTP ball bearing's inner race, outer race and bearing material are made from 440C steel. Each ball consists of 11 rolling elements with a ball diameter of 20.6 mm (0.81 in.). The bearing bore diameter is 60 mm (2.36 in.), the bearing outer diameter of 130 mm (5.12 in.) and the bearing inner and outer ring widths of 30.6 mm (1.2 in.). The HPOTP and HPFTP ball bearings have a diametrical clearance of 0.1698 mm (0.00668 in.) and a pitch diameter of 100 mm (3.94 in.). A prediction of the ball-race dry coefficient of friction is 0.25. The dry coefficient of friction between the ball and cage is predicted to be 0.2. Initially, the ball bearing in the HPOTP will be cooled with liquid oxygen (LOX) at -145°C (-220°F) while the HPFTP ball bearing will experience a -202°C (-331.6°F) liquid hydrogen (LH2) coolant.

The ATD HPOTP and HPFTP roller bearings have inner races made of 440C (AMS 5618) and outer races made of AISI 9310 (AMS 6265). The outer race material was chosen to reduce cracking possibilities which occurred when 440C was used for the outer race material. Both roller bearings contain 14 rolling elements. The HPOTP roller bearing has a roller length of 15 mm (0.59 in.) and roller diameter of 15 mm (0.59 in.). Its roller crown radius is 695.9 mm (27.4") and its roller flat length is 7.62 mm (0.3 in.). The HPOTP roller bearing has a bore diameter of 73 mm (2.87 in.), a bearing outer diameter of 133 mm (5.24 in.), an inner-ring width of 28.64 mm (1.1275 in.) and an outer ring width of 30.54 mm (1.2024 in.). The HPOTP roller diametrical clearance is -0.0381 mm (-0.0015 in.). The HPFTP roller bearing has a length of 17 mm (0.669 in.) and a roller diameter of 17 mm (0.669 in.). Its roller crown radius is 632.5 mm (24.9 in.) and its roller flat length is 8.89 mm (0.35 in.). The HPFTP roller bearing also has a bore diameter of 73 mm (2.87 in.)

ORIGINAL PAGE IS
OF POOR QUALITY

but a bearing outer diameter of 127 mm (5 in.), an inner ring width of 34.61 mm (1.3625 in.) and an outer ring width of 32.54 mm (1.281 in.). The HPFTP roller diametrical clearance is -0.0685 mm (-0.0027 in.) and a pitch diameter of 103 mm (4.06 in.). A prediction of the roller-race, roller-cage, and flange-end dry coefficient of friction for the HPOTP and HPFTP roller bearings is 0.1. Initially, the HPOTP roller bearing will be cooled with liquid hydrogen (LH2) at -118°C (-180.4°F); whereas, the HPFTP roller bearing will experience cooling from liquid hydrogen at -202°C (-331.6°F).

At the 109% full-power level, the HPOTP shaft speed or bearing inner ring speed is predicted to be 25,977 rpm. In contrast, the 65% power level for the HPOTP is approximately 17,000 rpm. For the HPFTP, its 109% power level shaft speed is 38,482 rpm while at 65% power level, its shaft speed is 28,500 rpm.

The dimensions for the inner and outer shaft diameters and the x-direction locations of these diameter changes were determined by measuring the distances from a blue line drawing of the ATD HPOTP and HPFTP. A scale factor of 1.05 was used to magnify the measured distances to their full-scale values.

The hydrodynamic loads and their x-direction locations locating their effect on the shaft-bearing system are shown for both the ATD HPOTP and HPFTP two-bearing models in the Appendices of this report. These loads which represent the radial hydrodynamic effect of the major components of the HPOTP and HPFTP on the shaft are represented by concentrated force vectors. The force vectors that are in a three dimensional space were resolved into a x-y and x-z plane component form for entry into the SHABERTH input file. Also, in the Appendices are two tables from Pratt and Whitney [10] predicting the hydrodynamic radial effects and bearing reaction forces for 65%, 90% and 109% power levels. For each table, an angle orientation based on LH2 and LOX inlet and outlet ducts are shown. This angular orientation is also shown schematically in the load x-location diagrams for the 109% power level that are also located in the Appendices.

The axial preload that was placed arbitrarily on the shaft as a thrust load, is caused by the preload spring force axially loading the ball bearings in both turbopumps. For the ATD HPOTP, an axial preload of 1000 lb (4448 N) is imposed; whereas, for the ATD HPFTP, an axial preload of 650 lb. (2891 N) is imposed. This effect has been placed arbitrarily at the shaft's left end in the schematic loads diagrams in the Appendices. In the input file, the axial preload force was placed between the ball and roller bearings directed toward the pump-end ball bearing.

All of the inputs discussed can be found in the report's Appendices in correctly formatted locations. In the Appendices, an example input listing for the ATD HPOTP and HPFTP two-bearing model is presented. For each bearing, the bearing geometry is listed. Then, the initial steady-state temperatures for each bearing are listed at 13 locations on the bearing. These locations are explained in reference [5]. Next, the shaft dimensions and x-locations are presented and finally, the x-direction locations, with their corresponding radial loads and axial preload along with the x-direction bearing locations for both the x-y and x-z planes are presented.

At the end of each input file, SRS Technologies has added character strings to provide the user with options. The first option is the usage of a subroutine to calculate axial preload conditions for the bearings. The procedure to use the option is outlined in reference [9]. In this simulation, this option was not used. The axial preload for the HPOTP and HPFTP was manually placed into the load portion of the input file. The next option is the usage of SHABERTH only or the SHABERTH/SINDA iteration scheme. SINDA is a finite difference equation solver which uses the frictional heat generation output from SHABERTH based on the input initial temperatures and calculates a temperature field at predetermined points of the bearing. The results of the SINDA localized at the 13 temperature nodes on the bearing are substituted atop the original temperatures in SHABERTH until a 2°C thermal convergence occurs. At this time, SINDA models for the ATD ball and roller bearings do not exist. SRS Technologies is beginning work on a grid generation program to subdivide the nodal areas automatically. This program should expediate the process of inputting the nodal information and conductances into a SINDA input file which is a long and laborous task. Also, information concerning the coolant's flowrate and fluid properties must be added to the SHABERTH input file. These properties can be found in references [6,7] for LOX and LH2. For this project, only the SHABERTH program was used to model the turbopumps.

RESULTS

Due to the length of time necessary to obtain the data and to construct the SHABERTH input files, an extensive parametric study was not feasible. The HPOTP and HPFTP case presented are only at the 109% and 65% power levels. At the 109% power level, the HPOTP shaft speed is 25,977 rpm while the HPFTP shaft speed is around 38,482 rpm. At the 65% power level, it's predicted that the HPOTP shaft speed is 17,000 rpm, whereas, the HPFTP shaft speed is 28,507 rpm.

Table 1 results show bearing reaction forces required to support the hydrodynamic radial loads and axial preloads that exist in both turbopumps. In Tables 1a and 1b, a comparison is made between the two-bearing SHABERTH model and Pratt and Whitney bearing reaction forces for the HPOTP. Pratt and Whitney's results are based on calculations using their predicted hydrodynamic radial loads. Table 1a and 1b results show close agreement in bearing reaction loads. A similar comparison can be made in Tables 1c and 1d for the HPFTP two bearing model reaction loads and the Pratt and Whitney results.

Even though there is good agreement in bearing reaction loads, no agreement could be found between the two bearing model and Pratt and Whitney's one bearing SHABERTH model for frictional heat generation. For the roller bearing under a 500 lb. radial load, Pratt and Whitney predicts the HPOTP frictional rates to be 191 W (watts) at 15,000 rpm and 320 W at 25,000 rpm. For their HPFTP model which reacts to a 2000 lb. radial load, their one bearing SHABERTH model predicts 923 W at 30,000 rpm and 1100 W at 35,000 rpm. In Tables 2a and 2b, the two bearing model results are compared to a one bearing model generated by the author to attempt an explanation for the discrepancy between Pratt and Whitney and the author's results. As shown in Tables 2a and 2b for the HPOTP at two different power levels, the ball bearing heat generation results closely agree for the one and two bearing models. However, the roller bearing heat generation rates are not as close numerically, but are similar in magnitude. However, neither of these models agree with Pratt and Whitney results. In Tables 2c and 2d for the HPFTP model, again, the ball bearing heat generation rates are comparable, however, the roller bearing rates are quite different. It seems that the one-bearing model underpredicts the two-bearing model by about one-half. This discrepancy may be explained due to a convergence error message that occurred in the one-bearing model. The default 15 iterations were not adequate for the solution subroutine. So, the results for my one-bearing model may not have totally converged. Again, neither of the models' results were comparable to Pratt and Whitney's one bearing model.

Finally, Table 3 shows the maximum contact Hertzian stresses for the inner and outer races for the HPOTP and HPFTP one and two bearing models. For the HPOTP one and two bearing model, Tables 3a and 3b show a good comparison of Hertz stress for the ball bearing; however, the roller bearing results are not as close numerically, but are the same order or magnitude. Tables 3c and 3d show a similar pattern for the HPFTP one and two bearing models. No results were obtained from Pratt and Whitney regarding Hertz stress. They use the A.B. Jones bearing program to calculate contact

stresses. In the initial design, a contact stress of around 300,000 psi (2068.4 MPa) is predicted to exist in the roller bearings. Again, my results tend to underpredict this value.

At this time, the cause for the agreement in reaction loads, but the discrepancies in frictional heat rates and Hertz stresses between Pratt and Whitney's results and the author's results are not known. The author recommends further effort be expended to verify the similarity of input data between the two bearing model and Pratt and Whitney models. More documentation of Pratt and Whitney's input and output for their SHABERTH and A.B. Jones models would be helpful to perform a better verification of results. If the comparison of inputs is exact, then, the author recommends the possibility of subtle differences that could exist between Pratt and Whitney's and NASA/MSFC SHABERTH programs.

Table 1a: HPOTP Bearing Reaction Loads
 Axial Preload: 4483N (1000 lb)
 109% power - (25,977 rpm)

Two Bearing Model					Pratt & Whitney Results		
	F_X	F_Y	F_Z	F_R	Angle	F_R	Angle
	(N)	(N)	(N)	(N) (lb)	(degrees)	(lb)	(degrees)
Ball Brg.	-4375	-667	-580	883.9 (198.7)	221	195	229
Roller Brg.	0	1640	-341	1675.1 (376.6)	101.7	375	102

Table 1b: HPOTP Bearing Reaction Loads
 Axial Preload: 4483N (1000 lb)
 65% power - (17,000 rpm)

Two Bearing Model					Pratt & Whitney Results		
	F _X	F _Y	F _Z	F _R	Angle	F _R	Angle
	(N)	(N)	(N)	(N) (lb)	(degrees)	(lb)	(degrees)
Ball Brg.	-4400	90.3	-177	198.7 (44.67)	152.9	45	152
Roller Brg.	0	1331	25.1	1331.2 (299.3)	88.92	300	90

Table 1c: HPFTP Bearing Reaction Loads
 Axial Preload: 2891N (650 lb)
 109% power - (38,482 rpm)

	Two Bearing Model					Pratt & Whitney Results	
	F_X (N)	F_Y (N)	F_Z (N)	F_R (N) (lb)	Angle (degrees)	F_R (lb)	Angle (degrees)
Ball Brg.	-2797	-1242	473	132.9 (298.8)	20.8	300	25
Roller Brg.	0	9902	2349	10,176.8 (2287.8)	166.6	2275	167

Table 1d: HPFTP Bearing Reaction Loads
 Axial Preload: 2891N (650 lb)
 65% power - (28,507 rpm)

	Two Bearing Model					Pratt & Whitney Results	
	F_X (N)	F_Y (N)	F_Z (N)	F_R (N) (lb)	Angle (degrees)	F_R (lb)	Angle (degrees)
Ball Brg.	-2831	-952	287	994.3 (223.5)	16.78	225	20
Roller Brg.	0	7262	1419	7399.3 (1663.5)	168.9	1660	169

Table 2a: HPOTP Frictional Heat Generation Rates (Watts)
 109% power level (25,977 rpm)
 (cage heat neglected)

	Two Bearing Model			One Bearing Model		
	Inner Race	Outer Race	Total	Inner Race	Outer Race	Total
Ball Brg.	2535	2477	5012	2564	2328	4892
Roller Brg.	17.5	25.1	42.6	12.3	20.3	32.5

Table 2b: HPOTP Frictional Heat Generation Rates (Watts)
 65% power level (17,000 rpm)
 (cage heat neglected)

	Two Bearing Model			One Bearing Model		
	Inner Race	Outer Race	Total	Inner Race	Outer Race	Total
Ball Brg.	1321	694	2015	1340	656	1996
Roller Brg.	12.1	13.7	25.8	9.66	11.5	21.16

Table 2c: HPFTP Frictional Heat Generation Rates (Watts)
 109% power level (38,482 rpm)
 (cage heat neglected)

	Two Bearing Model			One Bearing Model		
	Inner Race	Outer Race	Total	Inner Race	Outer Race	Total
Ball Brg.	2499	9515	12,014	2585	8766	11,351
Roller Brg.	98.6	163	261.6	36.9	85.3	122.2

Table 2d: HPFTP Frictional Heat Generation Rates (Watts)
 65% power level (28,507 rpm)
 (cage heat neglected)

	Two Bearing Model			One Bearing Model		
	Inner Race	Outer Race	Total	Inner Race	Outer Race	Total
Ball Brg.	1732	2821	4553	1793	2653	4446
Roller Brg.	77.2	97.1	174.3	34.7	51.4	86.1

Table 3a: HPOTP Maximum Hertzian Contact Stresses (N/mm^2)
109% power level (25,977 rpm)

	Two Bearing Model		One Bearing Model	
	Inner Race	Outer Race	Inner Race	Outer Race
Ball Brg.	1989.2	1649.7	1990.3	1643.2
Roller Brg.	1433.5	1440.7	1250.9	1308.6

Table 3b: HPOTP Maximum Hertzian Contact Stresses (N/mm^2)
65% power level (17,000 rpm)

	Two Bearing Model		One Bearing Model	
	Inner Race	Outer Race	Inner Race	Outer Race
Ball Brg.	1947.3	1412.4	1913.9	1398.5
Roller Brg.	1459.7	1358.8	1322.6	1237.3

Table 3c: HPFTP Maximum Hertzian Contact Stresses (N/mm^2)
109% power level (38,482 rpm)

	Two Bearing Model		One Bearing Model	
	Inner Race	Outer Race	Inner Race	Outer Race
Ball Brg.	1806.5	1955.3	1946.2	1955.9
Roller Brg.	2192.2	2136.2	1605.3	1754.9

Table 3d: HPFTP Maximum Hertzian Contact Stresses (N/mm^2)
65% power level (28,507 rpm)

	Two Bearing Model		One Bearing Model	
	Inner Race	Outer Race	Inner Race	Outer Race
Ball Brg.	1776.6	1660.7	1853.8	1667.5
Roller Brg.	2168.8	2008.3	1669.7	1635.5

CONCLUSIONS AND RECOMMENDATIONS

Based on my limited results, no relationship can be established at this time concerning close agreement in the bearing loads; but, discrepancies in the bearing frictional heat rates and Hertz stresses between the author's and Pratt and Whitney's models. In a brief parametric study, it was found that frictional heat generation seemed to be sensitive to changes in diametrical clearance of the bearing, the dry coefficient of friction and the rotational speed of the shaft or bearing inner ring. Hertz stress seemed to be sensitive to the loading applied to the shaft-bearing system. The discrepancies between my one and two bearing models exist due to a convergence problem in the one bearing model solution scheme. The error warnings allowed for calculations to be executed; however, suggested that these calculations may not be totally converged to the 0.01 limit within 15 iterations.

Based on this study, several recommendations for future research in this are as follows.

1. Further efforts are needed to establish commonality in the input parameters between Pratt and Whitney and the independent SHABERTH user. This is especially important when the design process dictates changes in the ATD HPOTP and HPFTP bearing configurations.
2. The establishment of SINDA models are needed for HPOTP and HPFTP roller and ball bearing. This will allow a possible thermally converged solution to occur as a result of the SHABERTH/SINDA iteration scheme.
3. The parametric investigation concerning the sensitivity of bearing parameter inputs based on their effect to selected SHABERTH outputs as bearing frictional heat generation, Hertz stress, clearances and bearing reaction forces should be performed.

Hopefully, from these recommendations, a useful numerical model of the ATD HPOTP and ATD HPFTP can be constructed. These models could become an important independent source of information when comparing its results to bearing tester data on the eventual ATD turbopump test data. Also, parametric studies using these models can provide a relatively economical means to predict possible problem areas in bearing performances. However, for SHABERTH's results to be a reliable predictor of bearing

performances, it must have reliable inputs based upon both experimental data and analytical formulation. As the ATD develops from the design and development stages into the construction and testing stages, SHABERTH coupled with SINDA can become an important evaluation tool for the bearing performance when subjected to the various power levels experienced in flight.

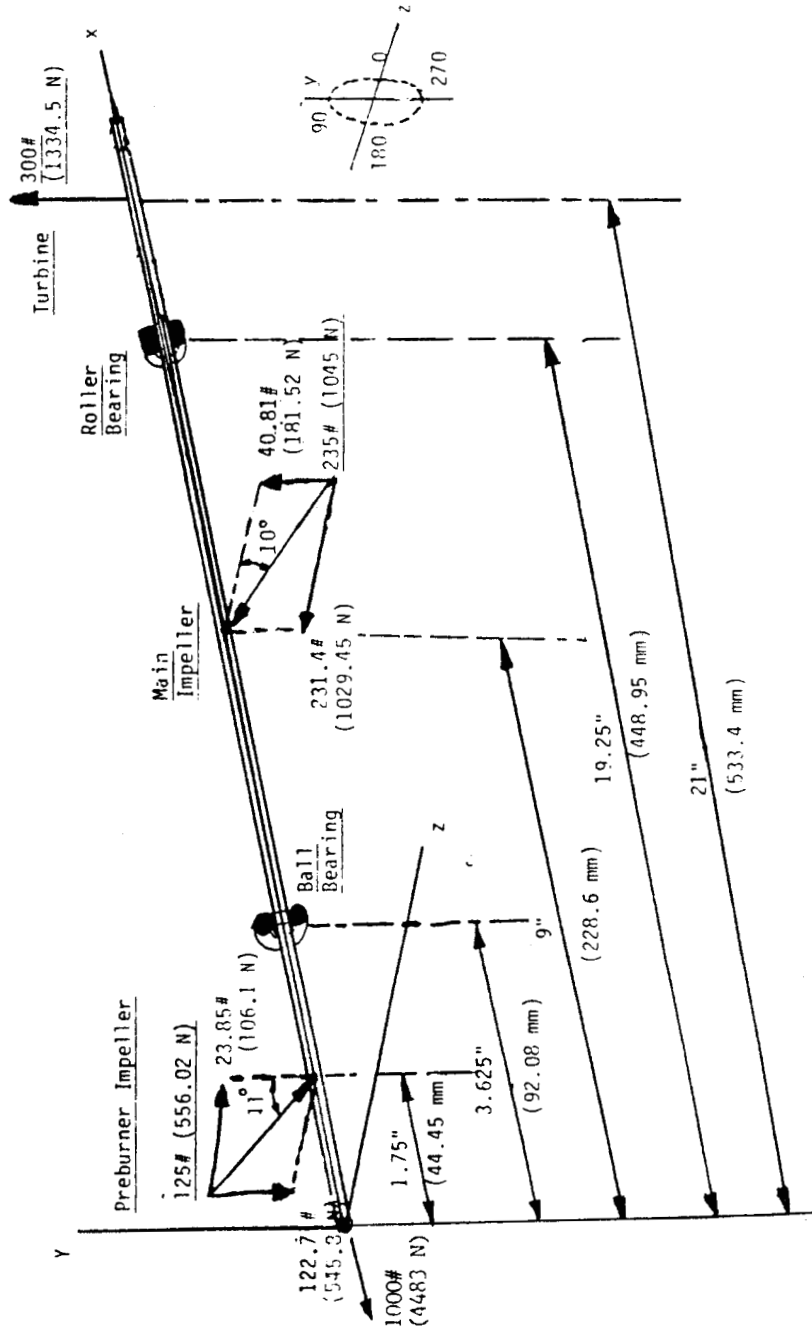
REFERENCES

1. Harris, Tedric A., Roller Bearing Analysis, 2nd edition. A Wiley-Interscience Publication of John Wiley and Sons, New York, NY, 1984.
2. Eschmann, Hasbargen, and Weigand, Ball and Roller Bearings; Theory, Design and Application. John Wiley and Sons, New York, NY, 1985.
3. Palmgren, Arvid, Ball and Roller Bearing Engineering, 3rd edition. SKF Industries, Inc., Philadelphia, PA, 1959.
4. Crecelius, W.J. and J. Pirvics, Computer Program Operation Manual on SHABERTH Computer Program for the Analysis of the Steady State and Transient Thermal Performance of the Shaft-Bearing Systems. Technical Report AFAPL-TR-76-90 (SKF report AL76PO30), Air Force Aero Propulsion Laboratory and Naval Air Propulsion Test Center, October, 1976.
5. Hadden, G.B., R.J. Kleckner, M.A. Ragen, L. Sheynin, Research Report-User's Manual for the Computer Program AT81Y003 SHABERTH (Steady-State, Transient Thermal, Ball, Cylindrical and Tapered Roller Bearings), SKF Report No. AT81D040, submitted to NASA-Lewis, Cleveland, OH under contract NAS3-22690, COSMIC program #LEW-12761, May 1981.
6. McCarty, R.D. and L.A. Weber, Thermophysical Properties of Oxygen from Freezing Liquid Line to 600°R for Pressures to 5000 psia. NBS-TN-384, U.S. Department of Commerce, National Bureau of Standards, July 1971.
7. McCarty, R.D. and L.A. Weber, Thermophysical Properties of Parahydrogen from the Freezing Liquid Line to 5000°R for Pressures to 10,000 psia, NBS-TN-617, U.S. Department of Commerce, National Bureau of Standards, April 1972.
8. SSME Alternate Turbopump Development Program Design Review Package (prepared under NASA contract NAS8-36801 for NASA/MSFC by Pratt and Whitney), FR-19821-1, May 16, 1987.

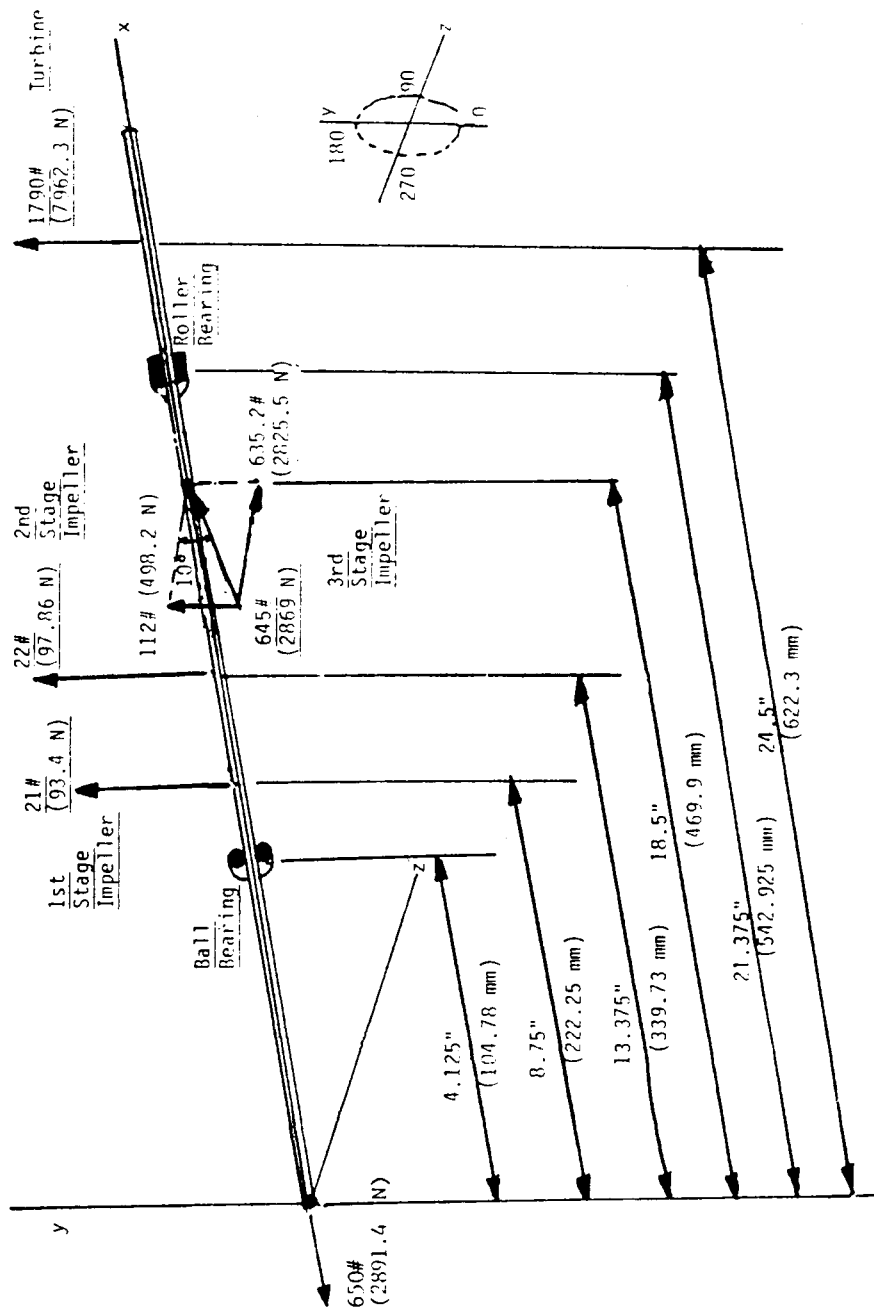
9. "User's Guide for the Mechanical/Thermal Model for the HPOTP Pump-end Bearings," SRS-Technologies, System Technology Division, Huntsville, AL, (not dated--written by D. Marty).
10. "SSME/ATD Side Load and Resultant Bearing Loads in Support of NASA Rid No. 109," memo from David B. Hudson to John L. Price, Pratt and Whitney Government Product Division Internal Correspondence, 88167E010047, June 15, 1988.

APPENDICES

Schematic of HPOTP Hydrodynamic Loads



Schematic of HPFTP Hydrodynamic Loads



ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

HPOTP Hydrodynamic Radial Loads [10]

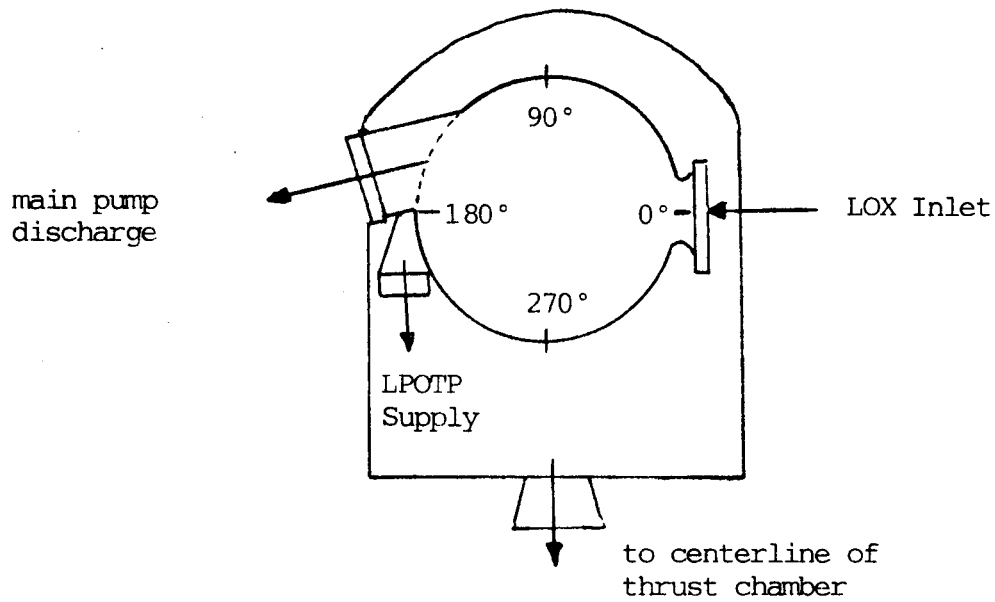
Device	Power Level					
	65%		90%		109%	
	Load (lb)	Angle (degree)	Load (lb)	Angle (degree)	Load (lb)	Angle (degree)
Preburner Impeller	87	67	106	84	125	101
Main-Stage Impeller	200	270	220	310	235	350
Turbine	200	270	260	270	300	270
Resultant Bearing Loads (lb)						
Ball Brg (pump-end)	45	152	115	160	195	229
Roller Brg. (turbine end)	300	90	365	97	375	102

HPFTP Hydrodynamic Radial Loads [10]

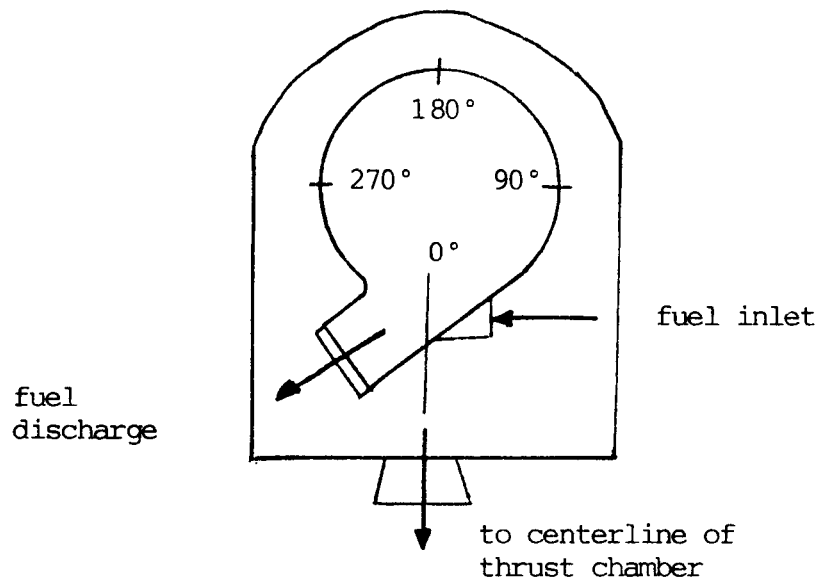
Device	Power Level					
	65%		90%		109%	
	Load (lb)	Angle (degree)	Load (lb)	Angle (degree)	Load (lb)	Angle (degree)
1st Stage Impeller	12	0	17	0	21	0
2nd Stage Impeller	13	0	18	0	22	0
3rd Stage Impeller	390	280	530	280	645	280
Turbine	1325	0	1580	0	1790	0
Resultant Brg. Loads (lb)						
Ball Brg. (pump end)	225	20	265	23	300	25
Roller Brg. (turbine end)	1660	169	2000	168	2275	167

ORIGINAL PAGE IS
OF POOR QUALITY

HPOTP (from pump-end looking toward turbine-end)



HPFTP (from pump-end looking toward turbine-end)



HPOTP Two-Bearing SHABERTH Input

SSME (P&W-ATD) LOX TURBOPUMP (ATD-HPOTP) -- BALL & ROLLER BEARINGS- 109%

25977. 2 0 5 -20 .001 .010 11

B1 440C 440C 3. 3. 0.

95.000 11 .1696 3.67 0

20.6000

.52 .58

.014 .014 .014 2. 2. 2.

+1 84.176 3.99 0.254 0.43 .136 3

.058420 -.11680 34.96 34.96 30.60 30.60

20.00 60.00 80.88 109.70 130.00 181.0

2.065E5 2.090E5 2.090E5 2.050E5 2.065E5

.279 .270 .270 .270 .275

8.210 7.740 7.740 7.740 8.210

11.16E-6 9.346E-6 9.346E-6 9.346E-6 10.980E-6

.0.250 .0.200

C2 440C (AMS 5618) AMS 6265 (AISI 9310) 1. 1. 0.

100.000 14 -.0381 0

15.000 15.000 695.96 7.62

15.000 15.000 20.

0.000 0.250 1.041 10.

0.200 0.200 0.20 1.500 1.50 1.50

0.200 0.200 0.20 1.500 1.50 1.50

-1.0 91.14 6.29 0.546 0.445 .1295

0.0508 -0.1168 28.64 28.64 30.54 30.54

20.00 73.0 85.0 115.0 127.0 178.0

2.188E5 2.120E5 2.120E5 2.180E5 2.231E5

0.294 0.262 0.262 0.296 0.280

8.220 7.250 7.250 8.190 7.750

10.850E-6 7.257E-6 7.257E-6 8.076E-6 12.522E-6

.0.100 .0.100

0.1

-145.-145.-145.-145.-145.-145.-145.-145.-145.-145.-145.-145.

-118.-118.-118.-118.-118.-118.-118.-118.-118.-118.-118.-118.

1 0.0 0.0 11.67 0.0 33.34 2.121E5

1 63.50 11.67 23.34 33.34 33.34 2.121E5

1 69.85 23.34 23.34 33.34 40.01 2.121E5

1 76.20 23.34 20.00 40.01 60.01 2.121E5

1 107.95 20.00 20.00 60.01 40.04 2.121E5

1 228.60 20.00 20.00 40.04 48.34 2.121E5

1 406.40 20.00 20.00 48.34 60.01 2.121E5

1 444.50 20.00 20.00 60.01 73.34 2.121E5

1 555.63 20.00 0.0 73.34 0.0 2.121E5

2 92.08 0.0 0.0

2 488.95 0.0 0.0

3 44.45 -545.8 0.0 0.0 0.0 0.0

3 228.60 181.5 0.0 0.0 0.0 -4448.

3 533.40 1334.5 0.0 0.0 0.0 0.

2 1 0.0 0.0

2 2 0.0 0.0

3 44.45 106.1 0.0 0.0 0.0 0.

3 228.60 -1029.5 0.0 0.0 0.0 0.

3

SBRGNUM

LOUT=1

SEND

SJSPRG

PREDEF=1.36839E-2,KSPRNG=82000., PLOAD=478.60, DELMX=15.5E-3,

INTCL=.0063, TCRUV=.06125

SEND

SFLGS

FLAG='NOSPRG', SFLAG='SHA', AXLOAD= .TRUE.,.TRUE.,.TRUE.,.TRUE.

SEND

Ball
Bearing

Roller
Bearing

Initial
Temperatures

Shaft
Dimensions

Radial &
Axial Loads
(x-y plane)

Radial Loads
(x-z plane)

SSME (P&W-ATD) LH2 TURBOPUMP (ATD-HPFTP) -- BALL & ROLLER BEARINGS- 109%

38482.	2	0	5	-20	.001	.010	1:
B1	440C		440C		3.	3.	0.
95.000		11	.1696	3.67		0	
20.6000							
.52	.58						
.014	.014	.014	2.	2.	2.		
+1	84.176	3.99	0.254	0.43	.136	3	
.058420	-.11680	34.96	34.96	30.60	30.60		
20.00	60.00	80.88	109.70	130.00	181.0		
2.065E5	2.090E5	2.090E5	2.050E5	2.065E5			<u>Ball</u>
.279	.270	.270	.270	.275			<u>Bearing</u>
8.210	7.740	7.740	7.740	8.210			
11.16E-6	9.346E-6	9.346E-6	9.346E-6	10.980E-6			
			0.250	0.200			
C2	440C(AMS 5618)	AMS 6265(AISI 9310)	1.	1.	0.		
103.000	14	-.0635		0			
17.000	17.000		632.50	8.89			
17.000	17.000				20.		
0.000	0.250		1.245	10.			
0.200	0.200	0.20	1.500	1.50	1.50		
0.200	0.200	0.20	1.500	1.50	1.50		<u>Roller</u>
-1.0	93.00	5.81	0.546	0.445	.1810		<u>Bearing</u>
0.1270	-0.1980	34.61	34.61	32.54	32.54		
20.00	73.0	86.0	120.0	133.0	178.0		
2.188E5	2.120E5	2.120E5	2.180E5	2.231E5			
0.294	0.262	0.262	0.296	0.280			
8.220	7.250	7.250	8.190	7.750			
10.850E-6	7.257E-6	7.257E-6	8.076E-6	12.522E-6			
			0.100	0.100			0.10
			1				
-202.-202.-202.-202.-202.-202.-202.-202.-202.-202.-202.-202.							<u>Initial</u>
-202.-202.-202.-202.-202.-202.-202.-202.-202.-202.-202.-202.							<u>Temperatures</u>
1	0.0	0.0	20.00	0.0	41.67	2.121E5	
1	69.85	20.00	20.00	41.67	60.00	2.121E5	
1	120.65	20.00	20.00	60.00	41.67	2.121E5	
1	198.40	20.00	20.00	41.67	48.30	2.121E5	
1	312.70	20.00	20.00	48.30	51.67	2.121E5	<u>Shaft</u>
1	436.60	20.00	20.00	51.67	58.34	2.121E5	<u>Dimensions</u>
1	501.65	20.00	20.00	58.34	63.34	2.121E5	
1	519.11	20.00	20.00	63.34	73.34	2.121E5	
1	657.23	20.00	0.0	73.34	0.0	2.121E5	
2	104.78	0.0		0.0			
2	542.93	0.0		0.0			
3	222.25	93.4	0.0	0.0	0.0	-2891.3	<u>Radial &</u>
3	339.73	97.9	0.0	0.0	0.0	0.	<u>Axial Loads</u>
3	469.90	498.2	0.0	0.0	0.0	0.	<u>(x-y plane)</u>
3	622.30	7962.3	0.0	0.0	0.0	0.	
2	1	0.0		0.0			
2	2	0.0		0.0			
3	469.90	2825.5	0.0	0.0	0.0	0.	<u>Radial Load</u>
							<u>(x-z plane)</u>
3							

SBRGNUM

LOUT=1

SEND

SJSPRG

PREDEF=1.36839E-2,KSPRNG=82000., PLOAD=478.60, DELMX=15.5E-3,
INTCL=.0063, TCRUV=.06125

SEND

SFLGS

FLAG='NOSPRG',SFLAG='SHA', AXLOAD= .TRUE.,.TRUE.,.TRUE.,.TRUE.

SEND

HPOTP One-Bearing SHABERTH Input (Ball Bearing)

```

SSME (P&W-ATD) LOX TURBOPUMP (ATD-HPOTP) - SINGLE BALL BEARING - 109%
25977. 1 1 1
B1 440C 440C 1. 1. 0.
95.000 11 .1696 3.67 0.
20.6
0.52 0.58
0.014 0.014 0.014 2.0 2.0 2.0
-1.0 84.18 3.99 0.254 0.432 .1360
0.250 0.200
-145.-145.-145.-145.-145.-145.-145.-145.-145.-145.-145.-145.
883.9 Radial Load -4448. Axial Load
Ball
Bearing
Initial
Temperatures

SERGNUM
LOUT=1
SEND
SJSPRG
PREDEF=1.36839E-2,KSPRNG=82000., PLOAD=478.60, DELMX=15.5E-3,
INTCL=.0063, TCRUV=.06125
SEND
SFLGS
FLAG='NOSPRG',SFLAG='SHA', AXLOAD=.TRUE.,.TRUE.,.TRUE.,.TRUE.
SEND

```

HPOTP One-Bearing SHABERTH Input (Roller Bearing)

```

SSME (P&W-ATD) LOX TURBOPUMP (ATD-HPOTP) -- SINGLE ROLLER BEARING - 109%
25977. 1 1 1
C1 440C(AHS 5618) AMS 6265(AISI 9310) 1. 1. 0.
100.000 14 -.0381 0.
15.000 15.000 695.96 7.62
15.000 15.000 20.
0.000 0.250 1.041 10.
0.200 0.200 0.20 1.500 1.50 1.50
0.200 0.200 0.20 1.500 1.50 1.50
-1.0 91.14 6.30 0.546 0.445 .1295
0.100 0.100 0.10
-118.-118.-118.-118.-118.-118.-118.-118.-118.-118.-118.-118.
1675.1 Radial Load
Roller
Bearing
Initial
Temperatures

SBRGNUM
LOUT=1
SEND
SJSPRG
PREDEF=1.36839E-2,KSPRNG=82000., PLOAD=478.60, DELMX=15.5E-3,
INTCL=.0063, TCRUV=.06125
SEND
SFLGS
FLAG='NOSPRG',SFLAG='SHA', AXLOAD=.TRUE.,.TRUE.,.TRUE.,.TRUE.
SEND

```

ORIGINAL PAGE IS
OF POOR QUALITY

HPFTP One-Bearing SHABERTH Input
(Ball Bearing)

```

SSME (P&W-ATD) LH2 TURBOPUMP (ATD-HPFTP) - SINGLE BALL BEARING - 109%
38482. 1 1 1
B1 440C 440C 1. 1. 0.
95.000 11 .1696 3.67 0.
20.6
0.52 0.58
0.014 0.014 0.014 2.0 2.0 2.0
-1.0 84.18 3.99 0.254 0.432 .1360
0.250 0.200
1
-202.-202.-202.-202.-202.-202.-202.-202.-202.-202.-202.-202.
1329.0 Radial Load -2891.3 Axial Load Initial
Temperatures

SBRGNUM
LOUT=1
SEND
SJSPRG
PREDEF=1.36839E-2,KSPRNG=82000., PLOAD=478.60, DELMX=15.5E-3,
INTCL=.0063, TCRUV=.06125
SEND
$FLGS
FLAG='NOSPRG' ,SFLAG='SHA', AXLOAD= .TRUE.,.TRUE.,.TRUE.,.TRUE.
SEND

```

HPFTP One-Bearing SHABERTH Input
(Roller Bearing)

```

SSME (P&W-ATD) LH2 TURBOPUMP (ATD-HPFTP) -- SINGLE ROLLER BEARING - 109%
38482. 1 1 1
C1 440C(AMS 5618) AMS 6265(AISI 9310) 1. 1. 0.
103.000 14 -.0635 0.
17.000 17.000 632.50 8.89
17.000 17.000 20.
0.000 0.250 1.245 10.
0.200 0.200 0.20 1.500 1.50 1.50
0.200 0.200 0.20 1.500 1.50 1.50
-1.0 93.00 5.81 0.546 0.445 .1810
0.100 0.100 0.10
1
-202.-202.-202.-202.-202.-202.-202.-202.-202.-202.-202.-202.
10176.8 Radial Load Initial
Temperatures

SBRGNUM
LOUT=1
SEND
SJSPRG
PREDEF=1.36839E-2,KSPRNG=82000., PLOAD=478.60, DELMX=15.5E-3,
INTCL=.0063, TCRUV=.06125
SEND
$FLGS
FLAG='NOSPRG' ,SFLAG='SHA', AXLOAD= .TRUE.,.TRUE.,.TRUE.,.TRUE.
SEND

```

N89-21747

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

FLUCTUATION ANALYSIS OF RELATIVISTIC NUCLEUS-NUCLEUS
COLLISIONS IN EMULSION CHAMBERS

Prepared by:	Stephen C. McGuire
Academic Rank:	Associate Professor
University and Department:	Alabama A&M University Department of Physics
NASA/ASEE:	
Laboratory:	Space Science
Division:	Astrophysics
Branch:	High Energy Astrophysics
NASA Colleague:	Thomas A. Parnell
Date:	August 15, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

FLUCTUATION ANALYSIS OF RELATIVISTIC NUCLEUS-NUCLEUS
COLLISIONS IN EMULSION CHAMBERS

by

Stephen C. McGuire
Associate Professor of Physics
Alabama A&M University
Normal, Alabama 35762

ABSTRACT

We have developed an analytical technique for identifying enhanced fluctuations in the angular distributions of secondary particles produced from relativistic nucleus-nucleus collisions. The method is applied under the assumption that the masses of the produced particles are small compared to their linear momenta. The importance of this work rests in the fact that enhanced fluctuations in the rapidity distributions is considered to be an experimental signal for the creation of the quark-gluon-plasma(QGP), a new state of nuclear matter predicted from quantum chromodynamical theory(QCD). In our approach, Monte Carlo simulations are employed that make use of a portable random number generator that allow the calculations to be performed on a desktop computer. The method is illustrated with data taken from high altitude emulsion exposures and is immediately applicable to similar data from accelerator-based emulsion exposures.

ACKNOWLEDGEMENTS

I am very grateful to Thomas A. Parnell and James H. Derrickson for serving as my colleagues and for providing valuable technical guidance during my appointment as a NASA/ASEE Summer Faculty Fellow. Many thanks go to Yoshi Takahashi for making data from the EMU05 experiment available to me for analysis. The administrative assistance and financial support of the NASA/ASEE Summer Faculty Fellowship Program, Michael L. Freeman, director and Ernestine Cothran, co-director, is gratefully acknowledged.

LIST OF FIGURES

- Figure 1. Cross sectional view of the emulsion chamber configuration designated as 5A2 for the EMU05 experiment showing its internal components.
- Figure 2. Coordinate system sketch showing the relationship between the linear momentum, p , of a secondary particle and its projection $p_{||}$, along the direction of the primary nucleus.
- Figure 3. Experimental pseudorapidity distribution for event 6-869. The solid curve is the result of a least squares fit of the data to the function:

$$f_s^4(\eta) = a_0/2 + \sum_{i=1}^4 a_k T_k(\eta).$$

where the T_k are Chebyshev polynomials of the first kind.

- Figure 4. Distribution of the V-fluctuations for a 1000 event simulation of 6-869. The smooth solid curve is a fit to a gaussian, $y(x) = a_1 \exp[(-1/2)\{(x - a_2)/a_3\}^2]$, where $a_1 = 100.2 \pm 0.49$, $a_2 = 318.7 \pm 0.28$ and $a_3 = 55.4 \pm 0.29$. The uncertainties in the parameters are purely statistical and correspond to one standard deviation.
- Figure 5. Distribution of the S-fluctuations for a 1000 event simulation of 6-869. The smooth solid curve is a fit to a gaussian, $y(x) = a_1 \exp[(-1/2)\{(x - a_2)/a_3\}^2]$, where $a_1 = 144.1 \pm 0.56$, $a_2 = 2800.0 \pm 0.38$ and $a_3 = 839.0 \pm 3.83$. The uncertainties in the parameters are purely statistical and correspond to one standard deviation.

INTRODUCTION

In recent years much interest has developed in the detailed study of collisions between nuclei at relativistic energies because it provides a microscopic laboratory for examining the basic structure of nuclear matter. This basic structure is perhaps best described in terms of the quark-model of matter. Quarks themselves are believed to be the most fundamental constituents of matter from which more complex particles are constructed. At normal nuclear densities they are bound together strongly in a given combination by other particles referred to as gluons. If, however, the energy density inside the nucleus can be raised to about 2-3 times its normal value, the theory of quantum chromodynamics₁ (QCD) predicts that nuclear matter will undergo a marked phase change¹. In the new phase, quarks and gluons are freed from each other to form a kind of ideal gas or plasma of particles currently referred to as the quark-gluon plasma (QGP). At even higher energy densities the plasma is expected to more closely approximate an ideal gas. The detailed properties of the QGP and the precise mechanisms(s) by which it is produced are not well understood at the present time and therefore merit thorough investigation.

As with most experimental efforts aimed at probing the structure of nuclear matter, we seek to determine the nature of the medium by careful examination of the properties of the particles produced by the collision. These physical quantities include the average transverse momentum, multiplicities, angular distribution structure and energy densities. Nucleus-nucleus collisions have an advantage over experiments involving lighter particles in that higher energy densities and larger particle multiplicities are produced on the average. Larger multiplicities mean that statistical fluctuations in the angular distributions should be smaller and correlated behavior should be more obvious². Indeed, one of the expected experimental signals for the creation of the QGP in a relativistic collision is the occurrence of enhanced fluctuations in the_{3,4} rapidity distributions of the secondary particles that are generated^{3,4}. Further, structures in the angular distributions of produced particles have been_{5,6} shown to contain evidences of collective flow of nuclear matter^{5,6}.

In this report we present a theoretical method for identifying nonstatistical fluctuations in the pseudorapidity distributions of individual events. An example of its application to data obtained from a high-altitude emulsion detector experiment is given. Further, accelerator-based experiments, involving the exposure of emulsion chambers to relativistic beams of oxygen and sulphur, are described. The latter hold the promise of providing data to help illuminate the physics of the system of colliding nuclei.

OBJECTIVES

The principal objective of this project is to analyze and interpret secondary particle distribution data, recorded in emulsion detectors, for the existence of nonstatistical structure. Attention is focussed on collisions between nuclei at relativistic and ultrarelativistic energies. Methods are to be developed that facilitate the reduction of data, test prevailing models of the collision process and suggest new ways of viewing data from accelerator-based experiments.

EXPERIMENT DESCRIPTION

Emulsion detectors that are used in this project are designed fundamentally to help determine the charge and paths of the particles that traverse them. The charges of the particles are identified partly from the ionization damage they produce in the CR39 plastic plates. The detector is operated inside of a uniform, constant magnetic field so that the curvature of the path indicates the sign of the charge and provides data on the momentum of the particle transverse to the beam direction.

A sketch of the experimental arrangement, showing the approximate dimensions of the chamber and its component parts is provided in figure 1. Each emulsion plate has a 70 μm base coated on both sides with 50 μm of emulsion. The separation between the plates is not a constant but gradually increases in the direction of the beam. This feature facilitates the measurement of the track curvature and the direction of deflection for the emitted particle. Lead plates are placed near the front of the detector where the density of emulsions plates is greater to increase the likelihood of interactions there. The larger density of emulsion plates also improves the accuracy with which the position of the collision vertex and emission angles can be determined. Low density, styrofoam spacers are used in the rear sections of the detector to maintain the separation between the emulsion plates.

Pulsed beams of ^{32}S ions at 200 GeV/nucleon were provided by the Super Proton Synchrotron (SPS) accelerator of the European Center for Nuclear Research. The duration of each pulse was 2s with a total intensity of 3×10^3 ions/cm²/pulse. The beam size was 2.54cm x 2.54cm and the chamber was exposed to beam spills shifted laterally from each other by 1 cm.

THEORY

Relationship Between the Rapidity and the Pseudorapidity:

We have chosen to express the distributions in terms of the pseudorapidity variable, a convenient form for which is derived below.

The value of the rapidity associated with a secondary particle is given by the equation

$$y = (1/2)\ln[(E + p_{||})/(E - p_{||})] \quad . \quad (1)$$

E is the total energy of the particle and $p_{||}$ is the component of its linear momentum parallel to the direction of the incident nucleus. Using the cartesian coordinate system shown in figure 2, y can be expressed as

$$y = (1/2)\ln\left\{\frac{\sqrt{p^2 + m^2} + p\cos\theta}{\sqrt{p^2 + m^2} - p\cos\theta}\right\} \quad , \quad (2)$$

where $E = \sqrt{p^2 + m^2}$ and $p_{||} = p\cos\theta$. In the limit of $m/p \ll 1$, equation (2) can be approximated by

$$y \approx (1/2)\ln[(1 + \cos\theta)/(1 - \cos\theta)] \quad , \quad (3)$$

where we now can make the definition of a new variable,

$$\eta = -\ln[(1 - \cos\theta)/(1 + \cos\theta)]^{1/2} \quad , \quad (4)$$

the pseudorapidity. Using the trigonometric half-angle identity,

$$\tan(\theta/2) = [(1 - \cos\theta)/(1 + \cos\theta)]^{1/2} \quad , \quad (5)$$

equation 4 simply becomes

$$\eta = -\ln[\tan(\theta/2)] \quad . \quad (6)$$

Since η does represent an approximation to y, care must be taken in choosing an appropriate bin size in constructing the distribution in η such that the character of the fluctuations will be preserved. Too large a bin width can have the effect of masking important structure, and too small a bin size can produce artificial variations in particle intensity. For practical purposes, the bin width, $\Delta\eta$, is chosen to match the resolution with which the values of the polar scattering angle can be determined. Use of η has the clear advantage that it is not necessary to know the mass of the secondary particle. It does,

however, rely on the accurate measurement of the laboratory scattering angles, θ_i , for the produced particles.

The values of θ_i are found from the cartesian coordinates of the tracks recorded by the emulsions. We have written a set of algorithms that perform this conversion using a conceptually simple procedure. Since the path of the particles are curved, an nth order polynomial is used to find the least squares fit to a each $[x_i, y_i]$ data set corresponding to a particular track. At present, the maximum value of n is 10. The polar emission angle, θ_i , is then estimated by calculating the tangent to the curve near the interaction vertex. The algorithms have been programmed in BASIC on a desktop computer and are written to be machine independent for portability.

Definition of the Fluctuation:

Two definitions of the fluctuation, given below, are used for this part of the analysis. In what follows L is the number of bins in the experimental distribution and $f(\eta_i)$ is the intensity in the ith bin.

$$\text{I.} \quad V_{\text{exp}} = \sum_{i=1}^{L+1} |f(\eta_i) - f(\eta)_{i-1}| \quad , \quad (7)$$

where $f(\eta_0) = f(\eta_{L+1}) = 0$.

$$\text{II.} \quad S_{\text{exp}} = \sum_{i=1}^L [f(\eta)_i - f_s(\eta)_i]^2 \quad . \quad (8)$$

Both of these expressions possess a minimum value of zero corresponding to the case where no fluctuations are present. Form I. is the total vertical length of the distribution and form II. gives a measure of the total deviation from a function f_s which describes the average features of the distribution. The function f_s is found by fitting the data to a superposition of Chebyshf polynomials⁸.

Using the form of the function, f_s , obtained from the fit, the experimental distribution for a given^s event is simulated a large number of times with a Monte Carlo procedure. For each simulation the number of secondary particles is held fixed to the observed value, and the fluctuations, given by equations 7 and 8 above, are computed and saved. Thus, a collection of values of V and S are generated for each set of simulations. When many simulations are used, the distributions

in V and S are expected to have a gaussian shape, in accordance with the central limit theorem.

APPLICATION OF METHODS AND RESULTS

In this section we give an example of the application of the theory, with results from a relativistic nucleus-nucleus collision. The data are from an event observed in emulsions that were flown at high atmospheric altitude in an experiment designed to study the heavy-ion component of the cosmic ray spectrum. The design¹⁰ and operation of the detector have been described in detail elsewhere¹¹. The event involved the interaction of an iron nucleus with the emulsion at 54 GeV/nucleon that produced 132 secondary particles. Initial interest in this event stemmed in part from an unusual bimodal structure in its particle intensity in the azimuthal direction¹¹. Prompted by the predictions of QCD, the possibility of related structure was anticipated in the polar direction. In figure 3, a histogram of the experimental pseudorapidity distribution for the event is plotted together with the function, f^4 , obtained by fitting the data to a polynomial given by

$$f^4 = v_0/2 + \sum_{k=1}^4 a_k T_k, \quad (9)$$

where T_k are Chebyshev polynomials of the first kind. The bin width was chosen to be 0.2 units of η . By randomly sampling f^4 , the experimental distribution was simulated 1000 times. A desktop personal computer, programmed in FORTRAN was used for this part of the calculation. The computer was equipped with a math coprocessor that reduced the time otherwise required for the calculations by a factor of 8 to 10. Central to the accuracy of the simulations is the reliability of the pseudorandom number generator. Its performance was evaluated by comparing the fluctuation distributions in V and S to an expected gaussian shape. The solid curve drawn on the histograms in figures 4 and 5 represents the result of a least squares fit of the data to a gaussian. The two plots are seen to be in good agreement with each other. Also shown on each figure is the location of the experimental value of the fluctuation. In each case it falls within the half-width of the theoretical curve, thus providing a strong indication that the intensity variations in the experimental distribution are likely due to statistical sources. The location of the experimental fluctuations within the probability distributions is by no means an expected result. It is, however, one which must be supported by predictions of reaction products based on deterministic models of the collision process. To this end, it is important that the analysis be extended to experiments that involve many central collisions of the same type that have higher numbers of produced particles and whose energy densities exceed the threshold values for QGP production.

CONCLUSIONS AND RECOMMENDATIONS

The methods described in this report places on a firm quantitative basis the identification of enhanced fluctuations in secondary particle distributions. The technique is general, in that it can be applied to any one-dimensional distribution. At present, the physical mechanisms for QGP production are not well understood. Therefore there continues to be a need for additional theoretical work that couple statistical analyses with deterministic modelling of the collision process. These investigations will rely mainly upon nucleus-nucleus collision data obtained at ground-based heavy-ion accelerators. It is expected, however, that information from accelerator experiments will complement data obtained from high-altitude emulsion detector exposures¹³.

One example of an event currently under examination is the collision, ³²S+Pb, at 200 GeV/nucleon. The collision was observed in the 5A2 emulsion detector during an exposure at CERN. Because of its energy and multiplicity ($N_s=399$), it is an excellent case for our analysis. Unfortunately, the process by which the emulsion plates are scanned to obtain the track coordinate data is not computerized, making the calculation of the individual secondary particle emission angles a time-consuming step of the process. An automated system for scanning the the emulsion plates, computerizing the coordinate data and converting it to emission angles is therefore a strongly recommended addition to this project.

REFERENCES

1. E. Shuryak, Quantum Chromodynamics and the Theory of Superdense Matter. Phys. Rep. 61C:71,1980.
2. L. McLerran and B. Svetitsky, American Scientist, Vol. 75, p. 490, 1987.
3. Physics Today, Vol. 4, No. 1, January, 1988, S-56.
4. M. D. Slaughter, LA-UR 88-617.
5. J. A. Zingman et al., Phys. Rev. C, Vol. 38, No. 2,(1988). p. 760.
6. P. L. Jain, K. Sengupta and G. Singh, Phys. Rev. C, Vol. 37, No. 2, 1988, p. 63.
7. Y. Takahashi et al., CERN Experiment Proposal, CERN/SPSC 85-50, SPSC/P216S.
8. M. Abramowitz and I. Stegun, Handbook of Mathematical Functions, NBS/APS No. 55, Nov. 1970, p. 795.
9. W. T. Eadie et al., Statistical Methods in Experimental Physics, North Holland, Amsterdam, 1971, p. 124.
10. T. H. Burnett et al., Phys. Rev. D, Vol 35, No. 3,(1987), p. 824.
11. W. V. Jones et al., Ann. Rev. Nucl. Sci., 1987, 37:71-95.
12. S. C. McGuire, NASA Contractor Report No. 4054, March,1987,pp. 10-16.
13. Y. Takahashi, Nucl. Phys., A478,(1988), 675c-683c.

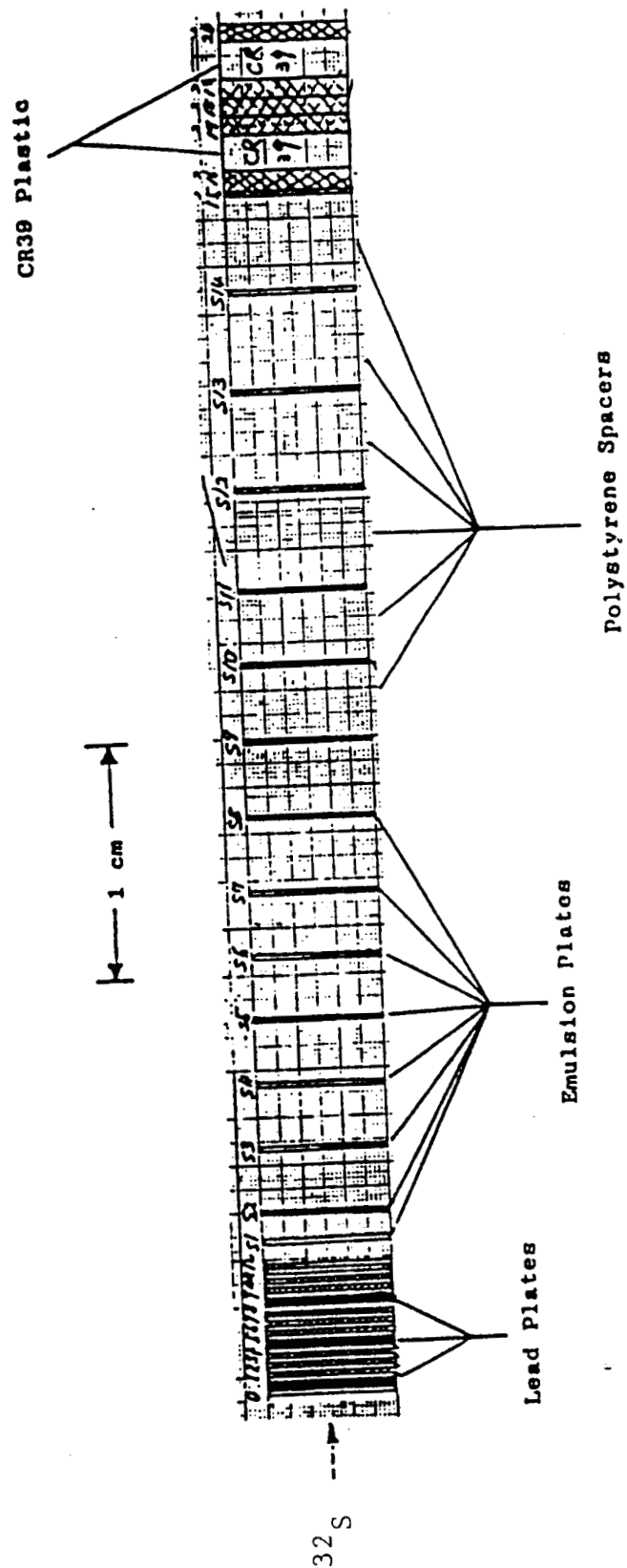


Figure 1.

ORIGINAL PAGE IS
OF POOR QUALITY

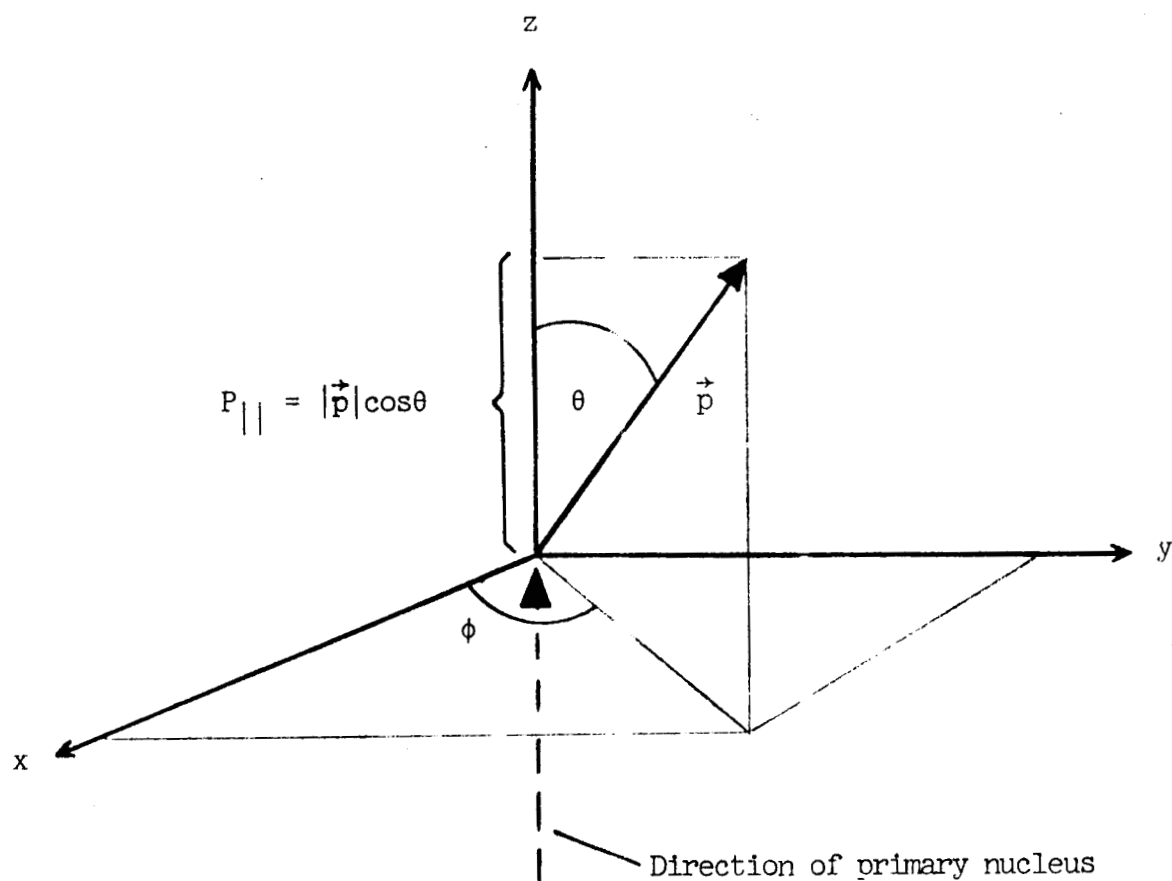


Figure 2.

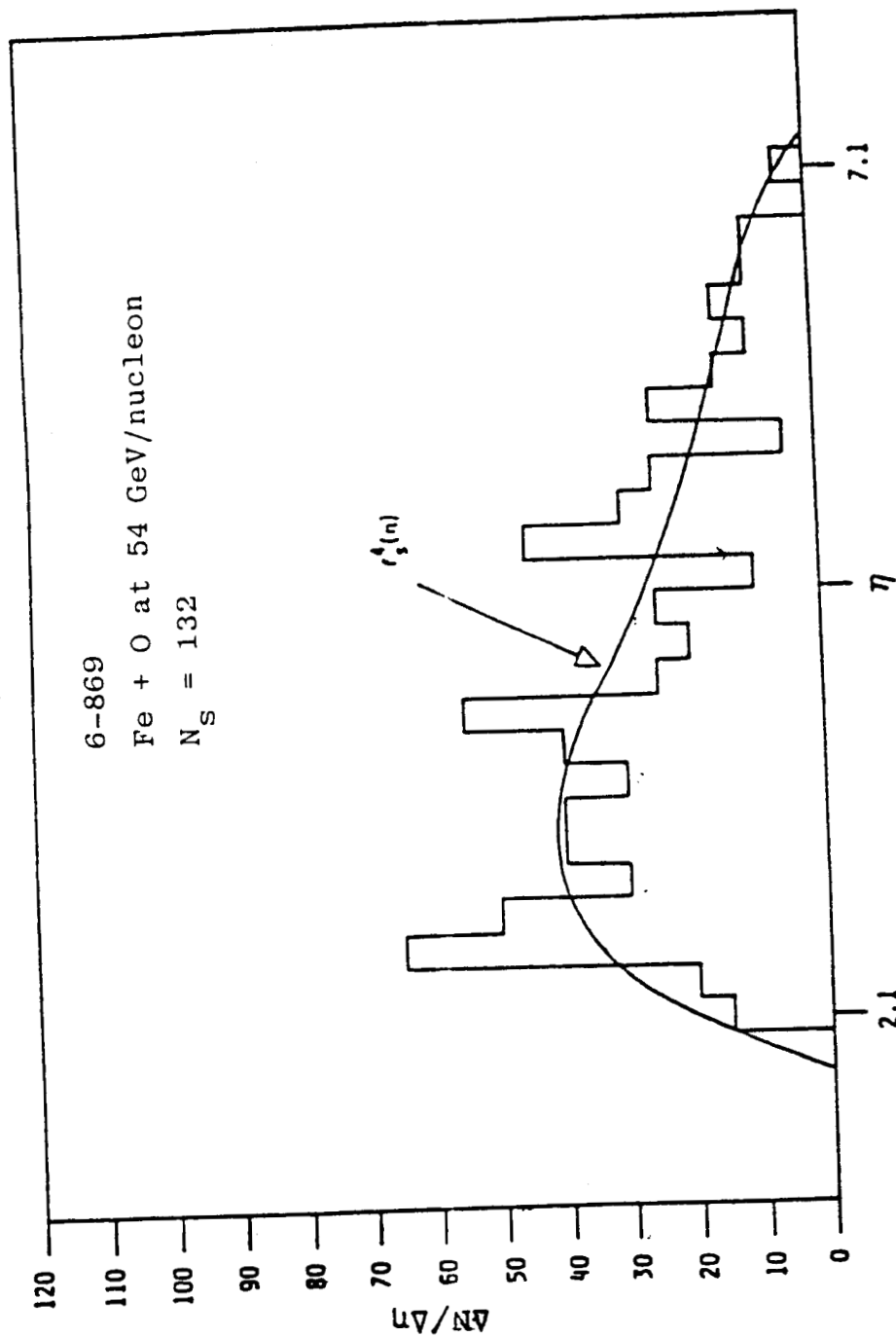


Figure 3.

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

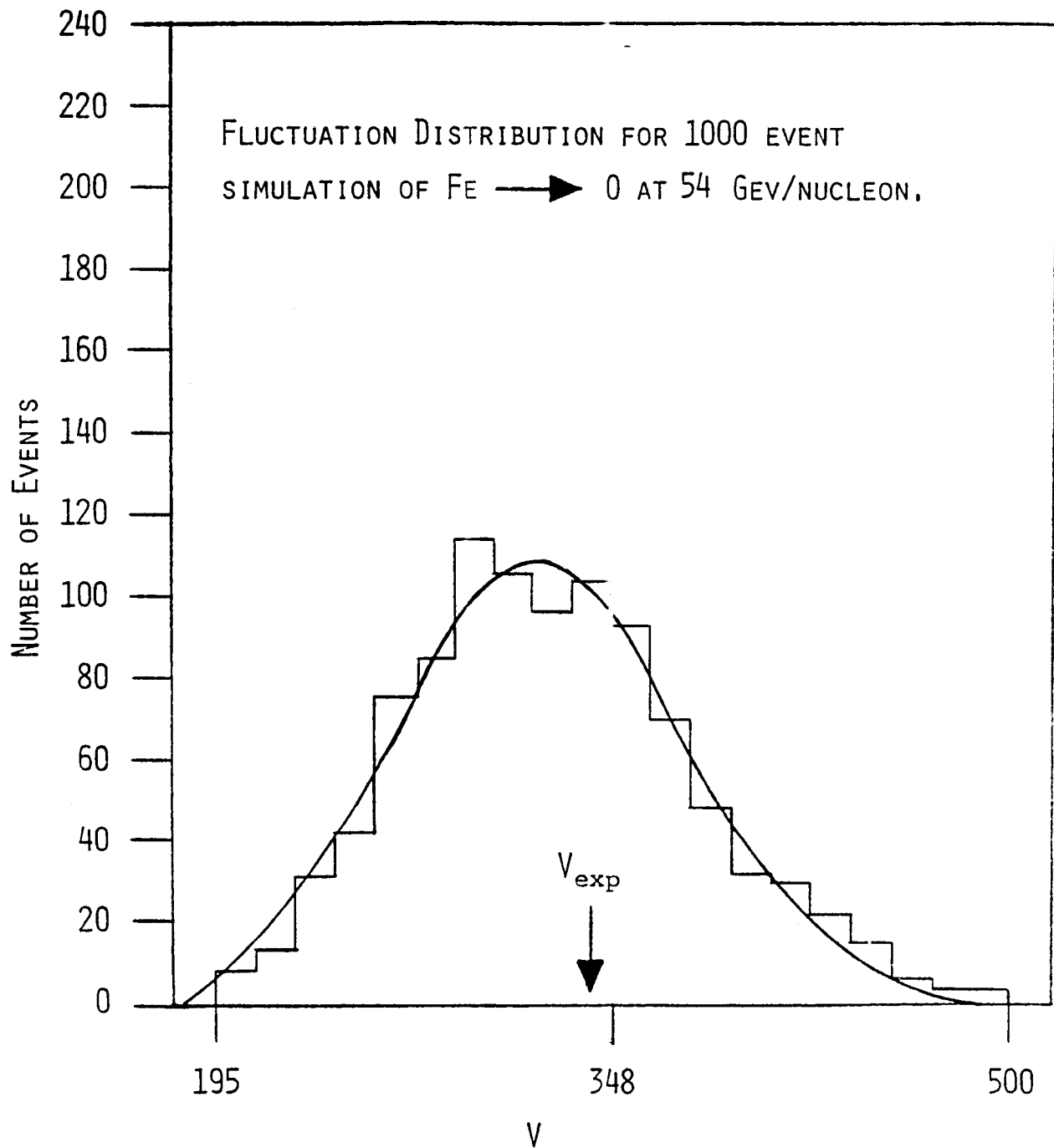


Figure 4.
XXI-11

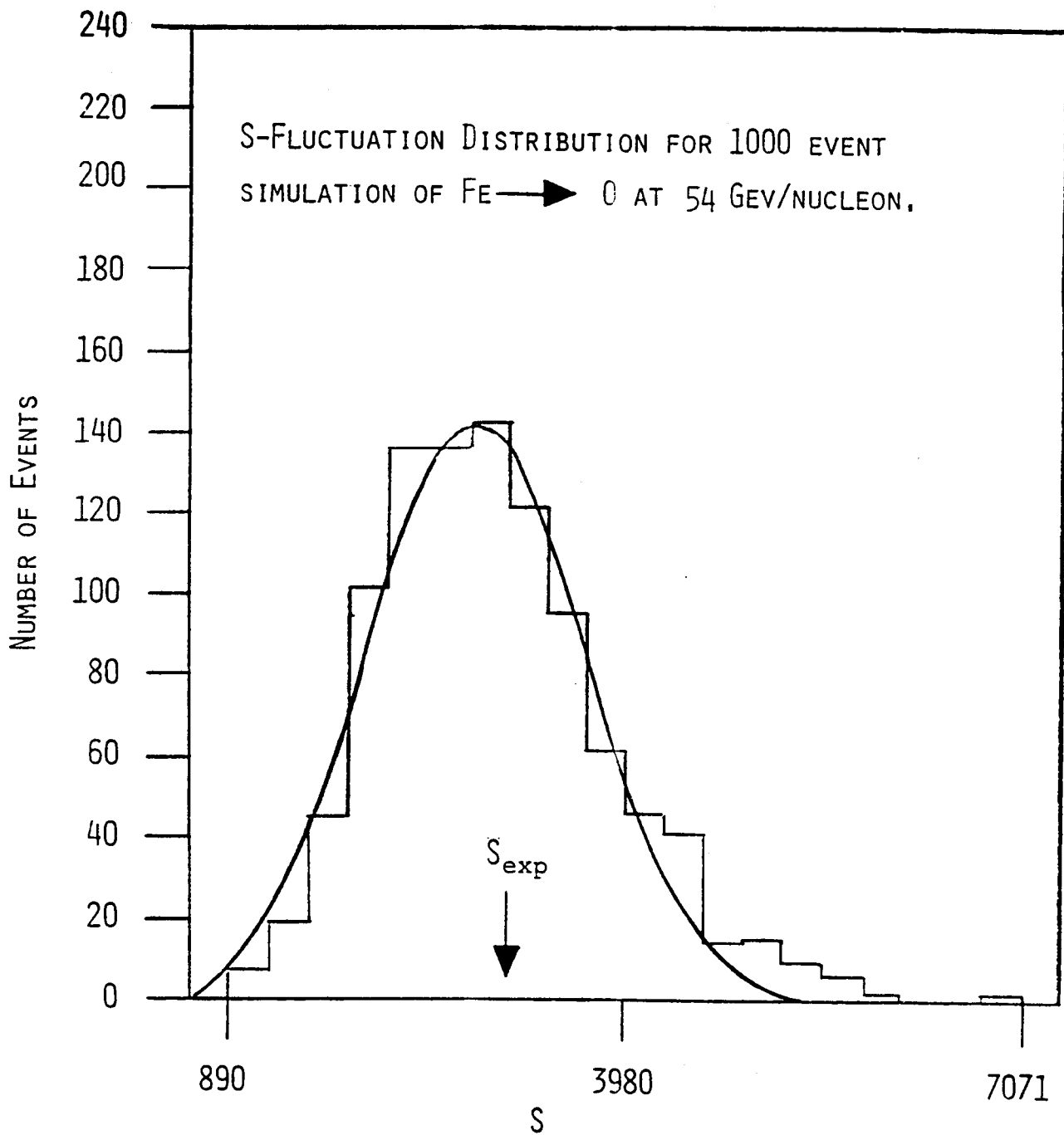


Figure 5.
XXI-12

N89-21748

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

ELECTRON MOBILITY IN MERCURY CADMIUM TELLURIDE

Prepared by:	James D. Patterson
Academic Rank:	Professor and Head
University and Department:	Florida Institute of Technology Physics/Space Science
NASA/MSFC:	
Laboratory:	Space Science
Division:	Low-Gravity Science
Branch:	Crystal Growth
NASA Colleague:	Sandor L. Lehoczky
Date:	August 19, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

ELECTRON MOBILITY IN MERCURY CADMIUM TELLURIDE

by

James D. Patterson
Professor and Head, Physics and Space Sciences
Florida Institute of Technology
Melbourne, Florida 32901

ABSTRACT

A previously developed program, which includes all electronic interactions thought to be important, does not correctly predict the value of electron mobility in mercury cadmium telluride particularly near room temperature. Part of the reason for this discrepancy is thought to be the way screening is handled. It seems likely that there are a number of contributors to errors in the calculation. The objective of this work is to survey the calculation, locate reasons for differences between experiment and calculation, and suggest improvements.

ACKNOWLEDGMENTS

The author wishes to thank the NASA/ASEE Summer Faculty Fellowship Program. Special thanks are due Dr. Mike Freeman, UA University Programs Co-Director and Ernestine Cothran, NASA MSFC Program Director.

Thanks are especially expressed to NASA Colleague Sandor L. Lehoczky and to Frank Szofran of the Space Science Laboratory.

The author is grateful to Shirley A. Buford, of the Crystal Growth Branch, for typing this manuscript.

Finally the author is grateful to the Administration of the Florida Institute of Technology for granting a leave of absence which allowed participation in this program.

1. Introduction

Many relevant properties of and references about $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ have already been given (Patterson (1987), see also Appendix 1).

The calculation of the transport properties of narrow band gap semiconductors (of which mercury cadmium telluride is an important example) have met with considerable success (Lehoczky et al. (1982), Szymanska et al. (1978), and Chattopadhyay and Nag (1974)). Nevertheless it is not unusual to find calculations (near room temperature) differing from experiment by a factor of order two.

The original idea of this project was that inadequate treatment of screening was the cause of discrepancies between calculation and experiment. However, in a calculation as complex as the mobility of electrons in a narrow band gap semiconductor (see Appendices 2 and 3) even the accuracy of the present calculations would have to be regarded as good (Lehoczky et al. (1980)).

An easy way to test the importance of screening is to run the present codes with different values of the screening constants. As we will see such calculations lead one to feel there is a limit to the accuracy with which the mobility can be evaluated, and certainly a limit to how much the accuracy can be improved by improving the screening.

2. Objectives

a. Determine the effect on the mobility of electrons in $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ of variations in relevant parameters. Pay particular attention to screening parameters.

b. Survey the theory on which the calculation is based to try to pinpoint areas that can be improved.

3. Calculations

We start by observing that in the approximation considered, the effects of screening are determined by the screening lengths. There exist reasons for considering other values of these lengths beside the conventional one (Patterson, (1987), appendices 3 and 5). Thus we are motivated to consider the results of calculations in which the screening length is varied for fixed values of the other parameters.

We also show how variations in screening affect the results for different values of the Cd concentration (x). All our results are based on a program of Dr. S.L. Lehoczky as rewritten by Ernestine Cothran. The program considers electron interaction effects due to longitudinal optic phonons, ionized impurities, holes, acoustic phonons, compositional disorder and neutral defects.

In what follows, we present our calculations in terms of hypotheses and calculational tests of the hypotheses.

Hypothesis

Although screening is important sizable changes in screening are not sufficient to get agreement with experiment.

TEST (μ is the mobility, k_{TF} represents all Fermi-Thomas wave vectors and T is the temperature).

$$f_E \equiv \frac{\mu(.5 k_{TF}^2, \chi=.183) - \mu_{exp}}{\mu(k_{TF}^2, \chi=.183) - \mu_{exp}},$$

at T = 300K, $n_D = 5.6 \times 10^{14} \text{ cm}^{-3}$ (the donor concentration)

$$f_E \approx .742,$$

at T = 100K, $n_D = 5.6 \times 10^{14} \text{ cm}^{-3}$

$$f_E \approx .665.$$

This is further shown by Figs. 1a and 1b where one sees that it is necessary (for the example shown) to use .05 of the nominal values for k_{TF}^2 in order to obtain near agreement with experiment.

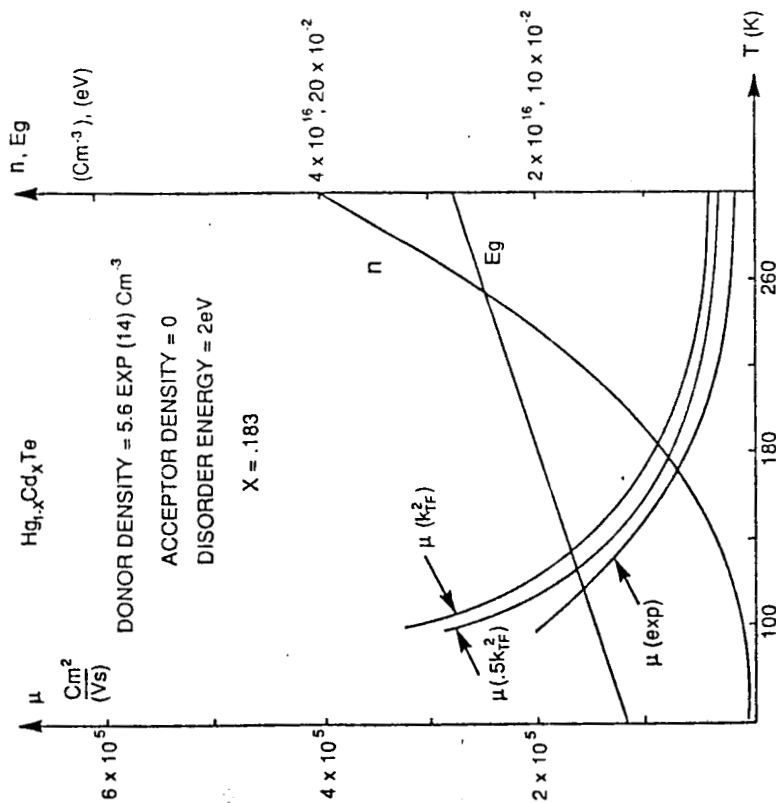


FIG. 1A

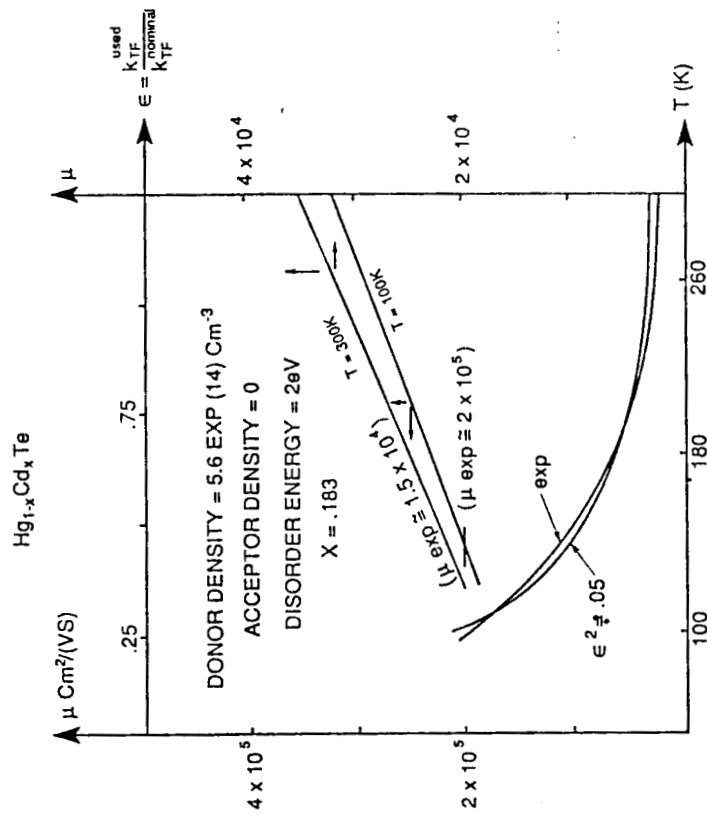


FIG. 1B

FIG. 1 (A) ELECTRON MOBILITY μ (cm^2/Vs) VS TEMPERATURE T (K), ALSO ENERGY GAP E_g AND ELECTRON CONCENTRATION N VS TEMPERATURE; (B) ELECTRON MOBILITY VS T AND SCREENING STRENGTH $\epsilon = k_{\text{IF}}^{\text{used}} / k_{\text{IF}}^{\text{nominal}}$.

Hypothesis

The mobility varies monotonically (very roughly linearly) as a function of inverse screening length (k_{TF}).

$$\epsilon^2 \equiv k_{TF}^{2 \text{ (used)}} / k_{TF}^{2 \text{ (nominal)}}$$

Test

at $x = .183$, $T = 100K$, $n_D = 5.6 \times 10^{14} \text{ cm}^{-3}$ ($\mu_{\text{exp}} \approx 200 \times 10^3 \frac{\text{cm}^2}{\text{Vs}}$)

ϵ^2	ϵ	μ	$\Delta\mu / \Delta\epsilon$
1	1	322,205	141,582
.75	.866	303,233	138,157
.5	.707	281,266	138,469
.25	.5	252,603	142,453
.2	.447	245,053	150,344
.1	.316	225,358	169,076
.05	.224	209,803	212,629
.01	.1	183,437	

at $x = .183$, $T = 300K$, $n_D = 5.6 \times 10^{14} \text{ cm}^{-3}$ ($\mu_{\text{exp}} \approx 15 \times 10^3 \frac{\text{cm}^2}{\text{Vs}}$)

ϵ^2	ϵ	μ	$\Delta\mu / \Delta\epsilon$
1	1	35,160	18,529
.75	.866	32,677	17,113
.5	.707	29,956	16,184
.25	.5	26,606	15,641
.2	.447	25,777	15,908
.1	.316	23,693	16,902
.05	.224	22,138	20,452
.01	.1	19,602	

Hypothesis

The more carriers, the more screening we should have. Thus, the more the calculation should depend on variations in the screening length. Three ways the number of carriers can be varied are by varying the Hg concentration, varying the temperature T or varying the donor concentration.

Tests (These imply the above hypothesis is naive.)

at $T = 300K$, $n_D = 5.6 \times 10^{14} \text{ cm}^{-3}$

$$S_x = \frac{\mu(k_{TF}^2, x=.183) - \mu(.5k_{TF}^2, x=.183)}{\mu(k_{TF}^2, x=1) - \mu(.5k_{TF}^2, x=1)} = 5.004$$

at $x = .183$, $n_D = 5.6 \times 10^{14} \text{ cm}^{-3}$

$$S_T = \frac{\mu(k_{TF}^2, T=300) - \mu(.5k_{TF}^2, T=300)}{\mu(k_{TF}^2, T=100) - \mu(.5k_{TF}^2, T=100)} = .127$$

at $x = .183$ $T=300K$ $n_D = 5.6 \times 10^{14} \text{ cm}^{-3}$

$$S_{n_D} = \frac{\mu(k_{TF}^2, 2n_D) - \mu(.5k_{TF}^2, 2n_D)}{\mu(k_{TF}^2, n_D) - \mu(.5k_{TF}^2, n_D)} = 1.004$$

Hypothesis

At a fixed temperature, changes due to screening are more important for smaller band gaps. This is shown clearly in the Figs. 2a and 2b.

A simple expression for the energy gap as a function of temperature and x is given by Dornhaus and Nimtz (1983) who also list more accurate expressions. The expression is only supposed to be used for $.135 \leq x \leq .203$. In the table below the energy gap is evaluated from a slightly better expression in the program.

$$E_g = -.3 + 5 \times 10^{-4} T + (1.91 - 10^{-3} T)x \text{ in eV.}$$

We find at $T = 300K$,

$$E_g = -.15 + 1.61x.$$

If x_0 is the value of x for which the energy gap vanishes, we find for $T = 100$ that $x_0 = .138$.

Define $f_s = \frac{\mu(.5 k_{TF}^2, x)}{\mu(k_{TF}^2, x)}$, $\Delta = \mu(k_{TF}^2, x) - \mu(.5 k_{TF}^2, x)$.

For $n_D = 5.6 \times 10^{14} \text{ cm}^{-3}$ and $T = 300K$

x	$E_g(\text{eV})$	$n_E (\text{cm}^{-3})$	$\mu(.5 k_{TF}^2)$	$\mu(k_{TF}^2)$	f_s	Δ	$\Delta/\mu(k_{TF}^2)$
.14	.085	7.85×10^{16}	41,971 cm^2/vs	54,805	.766	12,834	.234
.15	.098	6.79 "	39,160	49,705	.788	10,545	.212
.16	.110	5.82	36,247	44,930	.807	8,683	.193
.17	.123	4.94	33,425	40,407	.827	6,982	.173
.18	.136	4.15	30,732	36,312	.846	5,580	.154
.19	.149	3.47	28,212	32,620	.865	4,408	.135
.20	.162	2.88	25,884	29,370	.881	3,486	.119

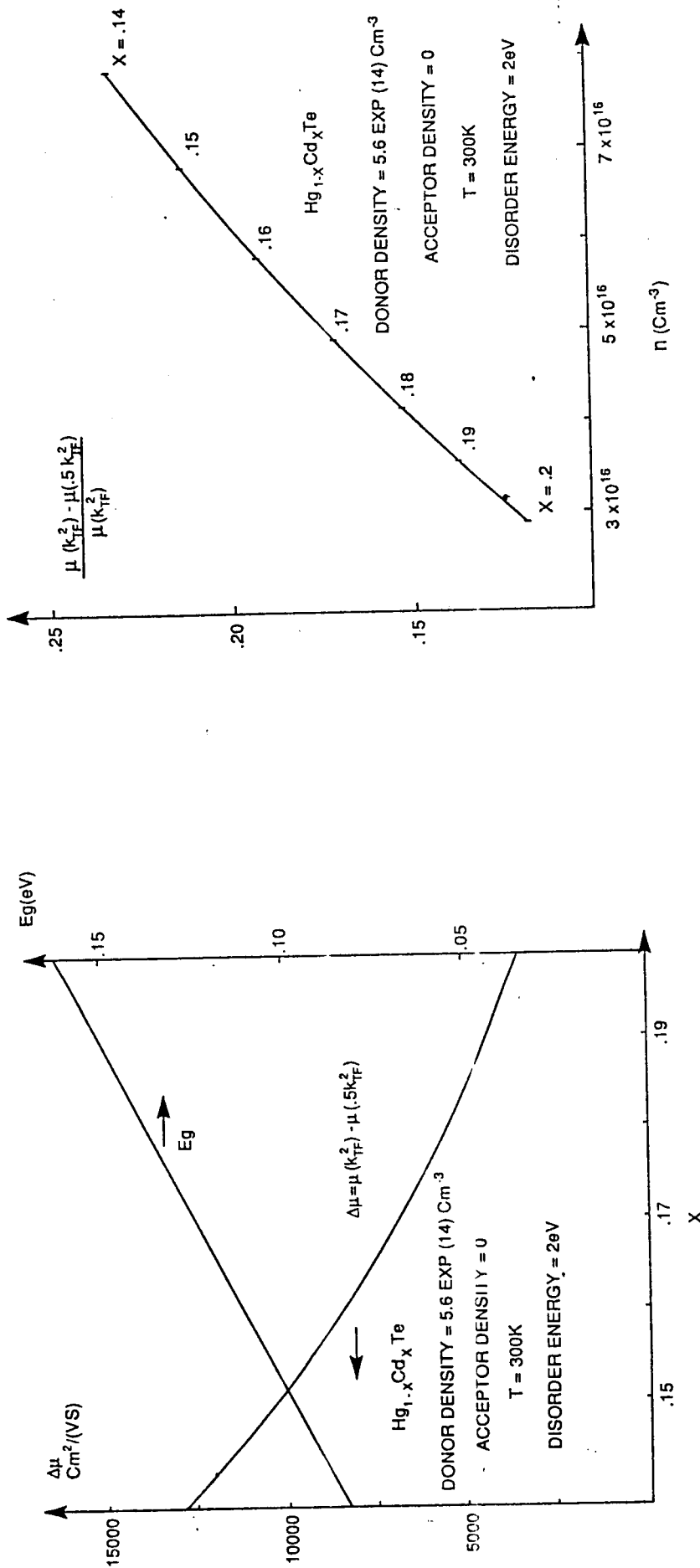


FIG. 2A

FIG. 2B

FIG 2. (A) ENERGY GAP E_g (eV) AND CHANGE IN MOBILITY $\Delta\mu$ (cm^2/Vs) ;
 (B) FRACTIONAL CHANGE IN MOBILITY VS ELECTRON CONCENTRATION.

4. Conclusions and Recommendations

a. Screening is important but it is not the whole answer. There are undoubtedly several small corrections.

b. In a calculation as complex as this one, see Figs. 3a and 3b and the appendices, the existing calculation (Lehoczky (1980), (1982)) is already good. See also Nag (1980) for comparison.

c. Small corrections and rewriting the code for a more modern computer are realistic goals for improvement. The program is presently implemented in basic for an HP 9845B and it runs rather slowly.

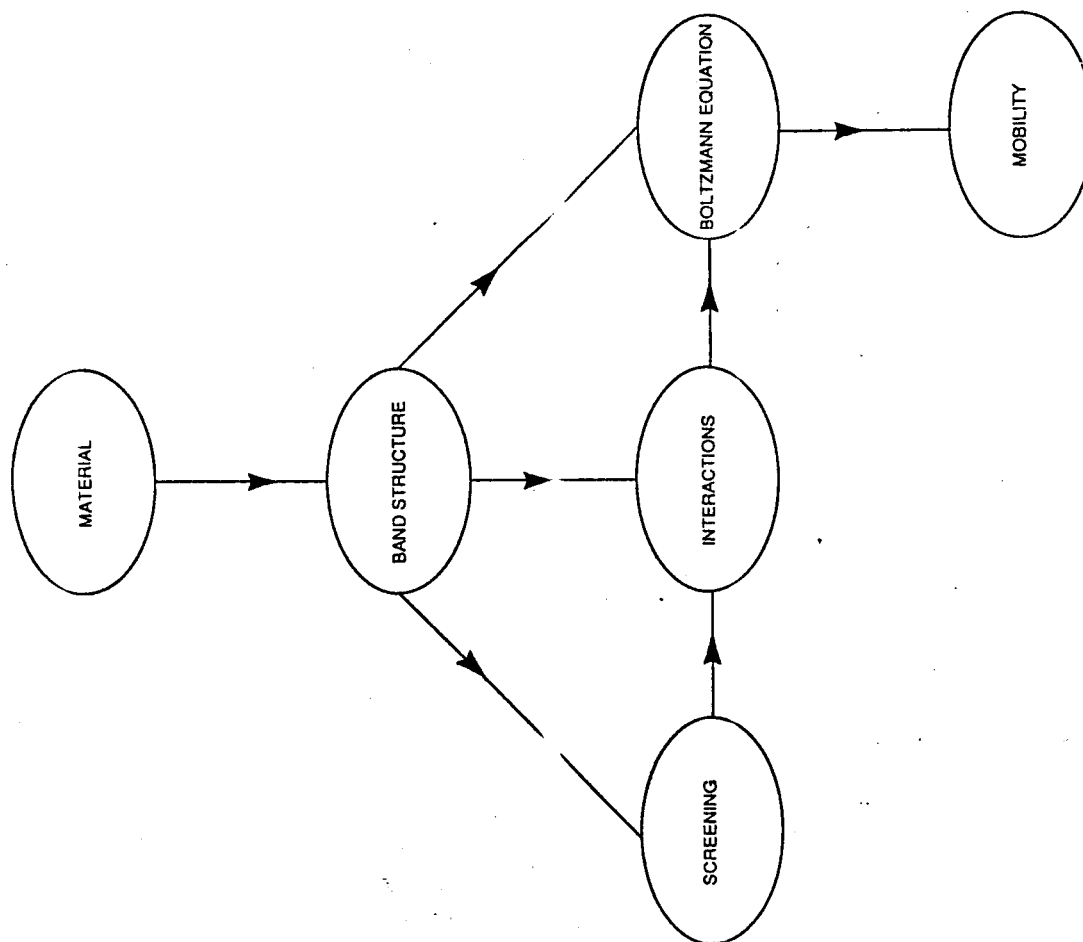


FIG. 3A

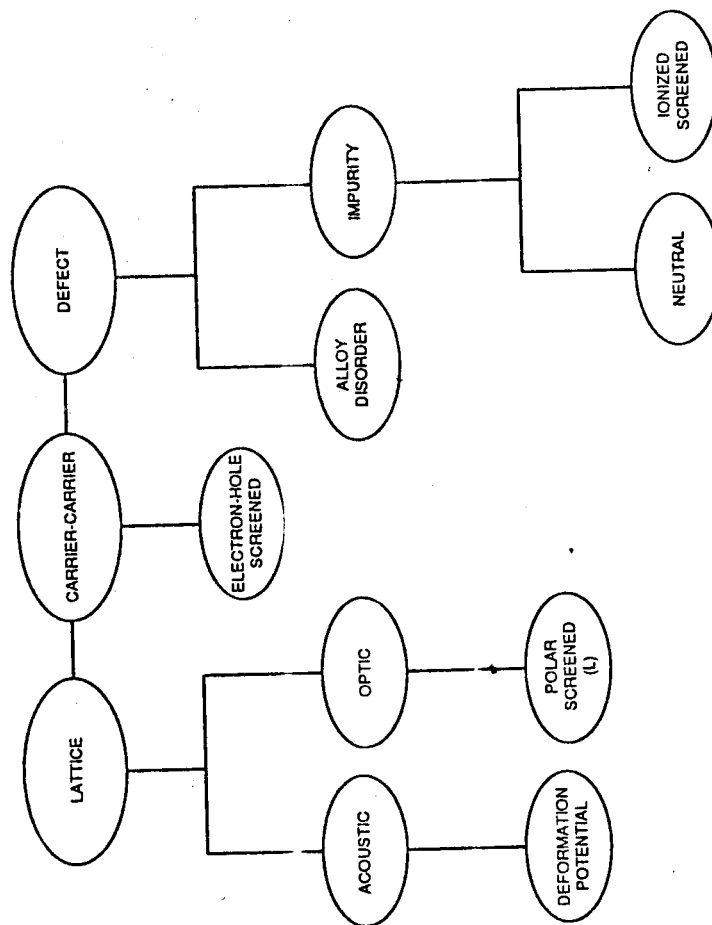


FIG. 3B

FIG. 3: THE CALCULATION; (A) OVERVIEW; (B) SCATTERING MECHANISMS.

REFERENCES

Books

1. Frank J. Blatt, "Physics of Electronic Conduction in Solids", McGraw-Hill Book Company, New York, 1968.
2. R. Dornhaus and G. Nimtz, "The Properties and Applications of the $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ Alloy System" in Springer Tracts in Modern Physics Vol. 98, Springer-Verlag, Berlin, 1983 p 119-300.
3. Charles J. Joachain, "Quantum Collision Theory", North-Holland Publ. Co. Amsterdam, 1975.
4. C. Kittel, "Introduction to Solid State Physics", John Wiley, 1986, 6th Edition.
5. Gerald D. Mahan, "Many-Particle Physics", Plenum, 1981.
6. B.R. Nag, "Electron Transport in Compound Semiconductor", Springer Series in Solid-State Sciences", Springer-Verlag, New York, 1980.
7. O. Madelung, "Introduction to Solid State Theory", Springer-Verlag (Berlin), 1978.
8. K. Seeger, "Semiconductor Physics", Springer-Verlag (Berlin) 1985, Third Edition.
9. A.H. Wilson, "The Theory of Metals", Cambridge U. Press, Cambridge, 1958.

Reports

1. S.L. Lehoczky, F.R. Szofran, and B.G. Martin, "Advanced Methods for Preparation and Characterization of Infrared Detector Materials, Part I", NASA CR-161598, July 5, 1980.
2. J.D. Patterson, "Improving the Method of Calculating Electronic Properties of Narrow Bandgap Semiconductors", NASA CR-179219, 1987, Marshall Space Flight Center pp XXVII ff.

Papers

1. J.M. Berroir and M. Voos, "Electronic Properties of II-VI Compound Superlattices", in "The Physics of the Two-Dimensional Electron Gas", Plenum Press, New York, 1987,
2. D. Chattopadhyay and B.R. Nag, "Mobility of Electrons in $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ ", J. Appl. Phys. 45, 1463 (1974)
3. D. Chattopadhyay and H.J. Queisser, "Electron Scattering by Ionized Impurities in Semiconductors", Rev. Modern Physics 53, 745 (1981).
4. J.J. Dubowski, T. Dietl, W. Szymanska, and R.R. Galazka, "Electron Scattering in $\text{Cd}_{1-x}\text{Hg}_x\text{Te}$ ", J. Phys. Chem. Solids 42, 351 (1981).
5. H. Ehrenreich, "Screening Effects in Polar Semiconductors", J. Phys. Chem. Solids 8, 130 (1959)
6. K. Jozwikowski and A. Rogalski, "Intrinsic Carrier Concentrations and Effective Masses in the Potential Infrared Detector Material $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$ ", Infrared Physics 28, 101 (1988).
7. Evan O. Kane, "Band Structure of Indium Antimonide", J. Phys. Chem. Solids 1, 249 (1957).
8. S.L. Lehoczky, J.G. Broerman, Donald A. Nelson, and Charles R. Whitsett, "Temperature-dependent electrical properties of HgSe", Phys. Rev. B 9, 1598 (1974).
9. S.L. Lehoczky, C.J. Summers, F.R. Szofran, and B.G. Martin, "Electrical Characterization of $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ Alloys", in Materials Processing in Reduced Gravity Environment of Space, Guy E. Rindone, Ed., Elsevier Science Publ. Co., Inc. Amsterdam, 1982.
10. D.A. Nelson, J.G. Broerman, C.J. Summers, and C.R. Whitsett, "Electron Transport in the $\text{Hg}_{1-x}\text{Cd}_x\text{Se}$ Alloy System", Phys. Rev. B 18, 1658 (1978).

11. T.F. Rosenbaum, Stuart B. Field, D.A. Nelson, and P.B. Littlewood, "Magnetic-Field-Induced Localization Transition in HgCdTe", Phys. Rev. Lett. 54, 241 (1985).
12. W. Szymanska and T. Dietl, "Electron Scattering and Transport Phenomena in Small-Gap Zinc-Blende Semiconductors", J. Phys. Chem 39, 1025 (1978).
13. W. Szymanska, "Electron Scattering and Transport Phenomena in Narrow Gap Mercury Compounds and Alloys", Physics of Narrow Gap Semiconductor, Ed. by J. Rauluszkiewicz, M. Gorska, and E. Kaczmanek, pp 357-379, Elsevier, Amsterdam, 1978.
14. W. Szymanska, "Electron Scattering and Classical Transport Phenomena in Small-Gap Semiconductors", Physics of Narrow Gap Semiconductors, Ed. by E. Gornik, H. Heinrich and L. Palmetshafer, Springer-Verlag, Berlin 1982, p. 45.
15. W. Zawadzki, "Electron Transport Phenomena in Small-Gap Semiconductors", Adv. Phys. 23, 435 (1974).

Appendix 1

Brief Properties of $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$:

1. Pseudo Binary II-VI Compound with structure isomorphic to zincblende.
2. The bandgap is tunable from 0 to 1.605 eV as x goes from about .17 (at low temperature) to 1.0. The bandgap also depends on temperature. $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ forms a continuous range of solid solutions between the semimetal HgTe and the semiconductor CdTe.
3. Can use as an infrared detector at liquid nitrogen temperature in the wavelength range 8-12 μ - an atmospheric window. A higher operating temperature than alternative materials, high detectivity, fast response, high sensitivity, IC compatible and low power.
4. Candidate for microgravity crystal growth.
5. Band structure involves mixing of unperturbed valence and conduction band wave functions as derived by the Kane theory. Non parabolic bands.
6. Typical carriers have small effective mass (about 10^{-2} \times free electron mass) which implies large mobility. The effective mass also has a strong dependence on concentration and temperature.
7. Polar scattering by LO lattice vibration modes dominate above 77K (or so) and the relaxation time approximation is not valid. The scattering is inelastic since the phonon energy is comparable to the electron energy.
8. Narrow bandgap semiconductors imply thermally excited electrons and hence screening. There may also be antiscreening effects when the phonons have higher phase velocity than the Fermi velocity of electrons.

Dornhaus and Nimtz (1983) list many more properties.

Appendix 2

Boltzmann Transport Equation and Solution

We just have a few comments to make here as the Boltzmann Transport equation is well reviewed in several texts (Blatt (1968), Wilson (1958)).

If f is the distribution function (with f_0 its value in equilibrium), E the energy, and if terms of order Φ^2 are neglected where

$$f = f_0 - \Phi(\vec{k}) \frac{\partial f_0}{\partial E}, \quad (A2.1)$$

then with electric fields \vec{E} and temperature (T) gradients in the x -direction

$$\begin{aligned} v_x \frac{\partial f_0}{\partial E} \left[|e| E_x + \frac{\partial E_F}{\partial x} + \left(\frac{E - E_F}{T} \right) \frac{\partial T}{\partial x} \right] \\ = \frac{1}{kT} \int V(\vec{k}, \vec{k}') [\Phi(\vec{k}) - \Phi(\vec{k}')] d\vec{k}', \end{aligned} \quad (A2.2)$$

where E_F is the Fermi energy v_x is the velocity, \vec{k} the wave vector, $|e|$ the magnitude of electronic charge, and $V(\vec{k}, \vec{k}')$ is determined by the electron scattering mechanism (it is proportional to the probability of scattering from \vec{k} to \vec{k}').

Since some of the scattering mechanisms are inelastic a variational principle is used to solve for $\Phi(\vec{k})$. The work is detailed, but well described ((Wilson (1958) Lehoczky (1980)). From f , the current density J_x (with $\frac{\partial T}{\partial x} = 0$) is

$$J_x = - \frac{|e|}{4\pi^3} \int f v_x d\vec{k}, \quad (A2.3)$$

and the mobility is $\mathcal{M} = J_x / (\eta |e| E_x)$, where η is the electron concentration.

Appendix 3

Electron Scattering Mechanisms

The details here are too long to write down. Fortunately they have already been well described (Lehoczky (1980), Lehoczky (1974), Nelson (1978), Dubowski (1981), Ehrenreich (1959), Zawadzki (1974), and Szymanska (1978)). Special mention should be made of Ehrenreich who wrote many of the key papers in this area. They are referenced in the above references.

a. Electron-phonon Interaction in Polar Solids

The only strong interactions are with longitudinal optic modes and we will limit discussion to long wave length modes near the center of the Brillouin zone with frequency ω_L . Several approximations then follow for which the literature may be consulted.

$$H_{el-ph} = \sum_{\vec{q}, \vec{k}, \sigma} \frac{M}{q \epsilon(\omega_L, q)} (a_{\vec{q}} + a_{-\vec{q}}^+) C_{\vec{k}+\vec{q}, \sigma}^+ C_{\vec{k}, \sigma} \quad (A3a.1)$$

The a's are the phonon annihilation and creation operators and the c's are the electron creation and annihilation operators. M is an appropriately chosen constant.

The screening is included by using the dielectric function

$$\epsilon(\omega_L, q) = \epsilon_{\infty} + \frac{(k_{FT}^e)^2 + (k_{FT}^{lh})^2}{q^2}, \quad (A3a.2)$$

and a "two mode" theory is necessary (with frequencies close to those of CdTe and HgTe). ϵ_{∞} is the high frequency dielectric constant.

Further Comment Screening and Antiscreening

Our main point here is to point out that antiscreening at small wave vectors can account for increased interaction and hence reduced scattering. This can occur when phonons with small wave vector also have high phase velocity which can occur for optical modes. As Ehrenreich (1959) has shown, the screened potential can actually be greater than the unscreened one.

b. Charged Impurity and the Friedel Sum Rule

This is also referred to as ionized impurity scattering (ii):

$$H_{ii} = \sum_{\vec{k}, \vec{q}} \sum_{\sigma} F_{\vec{k}+\vec{q}, \vec{k}} \frac{V(q)}{\xi(q)} C_{\vec{k}+\vec{q}, \sigma}^{\dagger} C_{\vec{k}, \sigma}, \quad (A3b.1)$$

where

$$F_{\vec{k}+\vec{q}, \vec{k}} = \frac{1}{\Omega} \int u_{\vec{k}+\vec{q}}^*(\vec{r}) u_{\vec{k}}(\vec{r}) dV, \quad (A3b.2)$$

with $u_{\vec{k}}(\vec{r})$ being the periodic modulation of the plane wave in Bloch's theorem and is often not included. $V(q)$ is the Fourier transform of the Coulomb potential. Screening is included by $\xi(q)$ and

$$\xi(q) = \xi_0 + \frac{1}{q} \left[(k_{FT}^e)^2 + (k_{FT}^{lh})^2 + (k_{FT}^{hh})^2 \right], \quad (A3b.3)$$

where ξ_0 is the low frequency dielectric constant.

Screening was more fully discussed in the first report (Patterson, 1987). Briefly, if a charge is placed in an electron gas its potential is better described by $e^{-\lambda r}/r$ than $1/r$ because the electrons in the gas surround the charge and "screen" it. The screening length is defined as λ^{-1} . The process is best described by a wave vector dependent dielectric constant, as above.

Further Comments: Multiple Scattering and Other Corrections to Calculations

Chattopadhyay and Queisser (1981) have discussed corrections due to multiple scattering and impurity dressing. The usual assumption with regard to ionized impurity scattering is that each impurity scatters independently. Multiple scattering adds corrections due to the electron interacting with more than one ion simultaneously. There is also a correction due to the fact that the impurities perturb the electron energies and wave functions. This is called the dressing correction and it can amount to 50% or so of the multiple scattering correction. However the corrections to ionized impurity scattering are not very important compared to other scattering mechanisms in $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ at temperatures of interest here. The dressing and multiple scattering corrections to ionized impurity scattering are more important at higher temperatures and lower impurity concentrations.

Corrections to the Born approximation (in a calculation of ionized impurity scattering) can also be important at lower temperature but second order Born corrections are probably only useful if they are small.

An expression for the screening length $L = \lambda^{-1}$ is given in appendix 5. It is possible to derive this directly or by use of the first Born approximation and the Friedel sum rule involving phase shifts δ_l (see e.g. Joachain, 1975). The Friedel sum rule holds generally and is supposed to be exact. It holds even for interacting electrons in a periodic system. The essential physical contents of the Friedel sum rule is the requirement of overall electrical neutrality about the charged ion. For electrons in a semiconductor, the Friedel sum rule takes the form, in the usual notation;

$$Z = (2/\pi) \sum_l (2l+1) \int \frac{d\delta_l}{dE} f(E) dE. \quad (A36.4)$$

c. Electron-Hole

When the heavy hole mass is large compared to the electron mass the hole can be considered to be stationary during scattering. This would make electron-hole scattering theory formally identical to ionized impurity scattering.

However we still must include non parabolity and admixture of p wave functions into the conduction band. For details see the references.

This is also a screened interaction. The following dielectric function has been used

$$\epsilon(q) = \epsilon_{\infty} + \frac{(k_{FT}^e)^2 + (k_{FT}^{eh})^2 + (k_{FT}^{hh})^2}{q^2} \quad (A3C.1)$$

d. Electron-Acoustic phonon

For semiconductors this is often treated by using the idea of a deformation potential (Seeger 1985, Madelung, 1978). If E_0 is the bottom of the conduction band then if E is the energy of the electron, we can obviously write

$$E = (E - E_0) + E_0, \quad (A3d.1)$$

where $(E - E_0)$ is treated as the kinetic energy and E_0 is viewed as the potential energy. One can think of E_0 varying with long wavelength displacements (\vec{J}) and this variation produces the deformation potential and thus the electron lattice coupling. At long wavelengths, Umklapp processes can be neglected. We write

$$H_{el-ph} = \frac{\partial E_0}{\partial V} \Delta V = \frac{\partial E_0}{\partial V} V \vec{\nabla} \cdot \vec{J} \propto \vec{J} \cdot \vec{g}, \quad (A3d.2)$$

for a single wave with wave vector \vec{g} . Further generalization leads to putting this in the usual form (see references). In the above V is the volume.

In the actual calculations, both longitudinal (L) and transverse (T) scattering is included by the use of three deformation potentials (1L, 2T).

e. Compositional Disorder and Neutral Defect
Compositional Disorder

This is calculated by considering a random distribution of scattering centers consisting of square wells.

The square wells are chosen to have a size equivalent to a unit cell and a depth (E_d) determined by the differences in band gaps between CdTe and HgTe.

If N_a is the number of unit cells per unit volume

$$\left. \frac{\partial f}{\partial t} \right)_{\text{dis. scatt}} \propto \frac{x(1-x) E_d^2}{N_d}, \quad (\text{A3c.1})$$

(see Lehoczky et al., (1980)).

The dependence on $x(1-x)$ is known as Nordheim's rule. For a given x , $x(1-x)$ measures the disorder assumed to be present (maximum).

Neutral Defect Scattering

This is also described as scattering by a random distribution of square-well potentials. It has been neglected in this calculation, but one assumes

$$\left. \frac{\partial f}{\partial t} \right)_{nd} \propto x(1-x) N E^2 V^2, \quad (\text{A3c.2})$$

(see Lehoczky et al., (1980)),

where N is the density of neutral defects, E is their strength and V is the volume of the cell.

Appendix 4

The Polaron Constant

The Polaron constant α is defined, in the usual notation, as

$$\alpha = \frac{e^2}{\hbar} \sqrt{\frac{m^*}{2\hbar\omega_L}} \left(\frac{1}{\epsilon(\infty)} - \frac{1}{\epsilon(0)} \right), \quad (A4.1)$$

It is a measure of the strength of the electron-lattice interaction and is dimensionless. For the phonon field and electrons to interact weakly we need $\alpha < 1$. For the Frohlich interaction term to be valid we also need $\alpha < 1$.

To get an estimate of α we use the following values:

$$\begin{aligned} \epsilon(\infty) &= 15, & |e| &= 4.8 \times 10^{-10} \text{ esu} \\ \epsilon(0) &= 20, & \hbar &= 1.05 \times 10^{-27} \text{ erg-s} \\ m^* &= .03m, & m &= 9.1 \times 10^{-28} \text{ gm} \\ \hbar\omega_T &= 2.4 \times 10^{-14} \text{ ergs} \end{aligned}$$

Thus

$$\alpha \cong .092.$$

Appendix 5

Band Structure and Effective Masses

We briefly summarize Kane (1957) theory here. We give some idea of the band structure near the top of the valence band and bottom of the conduction band. A simple schematic diagram of the band structure has been given by Seeger (1985), see also Lehoczky (1974). The s wave functions of the conduction band are found to mix with the p wave functions of the valence band by $\vec{k} \cdot \vec{p}$ perturbation theory (for $\vec{k} \neq 0$).

By using Bloch's theorem the wave function is given by

$$\psi_{n\vec{k}}(\vec{r}) = u_{n\vec{k}}(\vec{r}) e^{i\vec{k} \cdot \vec{r}}, \quad (A5.1)$$

in the usual notation. The Schrodinger wave equation can be written:

$$(H_0 + H') u_{n\vec{k}}(\vec{r}) = E_n' u_{n\vec{k}}(\vec{r}), \quad (A5.2)$$

where

$$H_0 = \frac{p^2}{2m} + V(\vec{r}) \quad (A5.3)$$

$$H' = \frac{\hbar}{m} \vec{k} \cdot \vec{p} + \frac{\hbar^2}{4m^2 c^2} (\vec{\nabla} V \times \vec{p}) \cdot \vec{\sigma}, \quad (A5.4)$$

and a $\vec{\nabla} V \times \vec{k}$ term has been neglected.

$$E_n' = E_n - \hbar^2 k^2 / (2m). \quad (A5.5)$$

The Pauli operation is $\vec{\sigma}$.

To solve for the required energies, the eigenvalues of an 8 x 8 matrix must be determined.

In Kane's notation we define:

$$P = -i \left(\frac{\hbar}{m} \right) \langle S | p_z | Z \rangle, \quad (A5.6)$$

and

$$\Delta \equiv \frac{3\hbar^2}{4m^2c^2} \langle X | \frac{\partial V}{\partial x} p_y - \frac{\partial V}{\partial y} p_x | Y \rangle, \quad (A5.7)$$

where the matrix elements involve s and p ($|X\rangle, |Y\rangle, |Z\rangle$) wave functions. If $\Delta \gg kP$, as seems to be appropriate for narrow band gap semiconductors, we get for the energy as measured from the bottom of the conduction band (neglecting a small term $\hbar^2 k^2 / 2m$):

$$E \equiv -E_g/2 + \sqrt{(E_g/2)^2 + \frac{2}{3} k^2 P^2}, \quad (A5.8)$$

where E_g is the energy gap between the conduction and the valence band.

Depending on the usage, various effective masses can be defined for spherical energy surfaces as above. We assume spherical energy surfaces and consistent with above (Chattopadhyay and Nag, (1974)):

$$E(1 + E/E_g) = \hbar^2 k^2 / 2m_0, \quad (A5.9)$$

where m_0 is the mass at the bottom of the conduction band

$$(m_0 = \frac{3\hbar^2 E_g}{4m P^2}).$$

Energy Effective Mass 1,

$$m_{E1} = \frac{\hbar^2 k^2}{2E}, \quad (A5.10)$$

Energy Effective Mass 2,

$$m_{E2} = \hbar^2 \left(\frac{d^2 E}{d k^2} \right)^{-1}. \quad (A5.11)$$

Momentum Effective Mass,

$$m_p = \hbar^2 k \left(\frac{dE}{dk} \right)^{-1}. \quad (A5.12)$$

($m_p \vec{v} = \hbar \vec{k}$ and \vec{v} and \vec{k} are parallel).
It is relatively easy to show

$$m_{E1}/m_0 = 1 + E/E_g, \quad (A5.13)$$

$$m_p/m_0 = 1 + 2E/E_g, \quad (A5.14)$$

$$m_{E2}/m_0 = (1 + 2E/E_g)^2 (1 - 2E/E_g)^{-1}, \quad (A5.15)$$

all three of these agree at the bottom of the conduction band
($E=0, k=0$).

Which one of these is appropriate depends on the usage.

The conventional effective mass as defined, for example, by Kittel (1986, p 194) is m_{E2} . The effective mass equation may be used as a starting point for discussing scattering in the Born approximation in solids. It would appear from a development of that equation (Madelung (1978), p. 79) that energy effective mass 1 could be the appropriate effective mass to use. For HgCdTe, m_p is typically about 10^{-2} of the mass of free electrons and has a strong dependence on carrier concentration and temperature.

By a standard development in the first Born approximation, the screening length comes out (see e.g. Patterson (1987)),

$$L_{m_{E1}}^{-2} = \frac{4e^2}{\pi \hbar^2} \int \left(-\frac{\partial f}{\partial E} \right) m_{E1} k dE. \quad (A5.16)$$

However, for this result to be applicable to non parabolic energy bands, the effective mass should be the momentum effective mass so that the density of states ($\rho(E)$) properly enters via

$$\rho(E) dE = \frac{1}{\pi^2} k^2 \frac{dk}{dE} dE \equiv \frac{m_p k}{\pi^2 \hbar^2} dE. \quad (A5.17)$$

Since $m_{E1} = \frac{1}{2} (m_p + m_0)$, (see above)

we find:

$$L_{m_{E1}}^{-2} = \frac{1}{2} (L_{m_p}^{-2} + L_{m_0}^{-2}) \quad (A5.18)$$

Then since

$$m_p > m_0 ,$$

$$L_{m_0}^{-2} \leq L_{m_p}^{-2} ,$$

we have

$$L_{m_{E1}}^{-2} \leq L_{m_p}^{-2} \quad (A5.19)$$

Hence a crude approach using the Born approximation could lead to a larger screening length, hence more scattering, hence less mobility.

Further Comment:

We also need to point out the usefulness of $\text{Hg}_{1-x}\text{Cd}_x$ as an infrared (IR) detector. Since

$$E = \hbar\omega = hc/\lambda ,$$

it is easy to show that

$$E_g \text{ (eV)} = 1.24/\lambda(\mu) . \quad (A5.20)$$

An important atmospheric window for IR is in the wavelength range 8-12 μ . At T=77K the energy gap is approximately (Dornhaus and Nimtz, 1983)

$$E_g = -.26 + 1.83x . \quad (A5.21)$$

We find that $.16 \leq x \leq .19$ will put us in the 8-12 μ window. There are other possibilities.

Appendix 6

Miscellany

a. Wigner Lattice

In an electron gas in which the kinetic energy is much less important than the potential energy, a localization or "crystallization" of the electrons can occur. It appears that this phenomena actually occurs at $x = .24$ in $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ at millikelvin temperature in a magnetic field (Rosenbaum et al. (1985)).

b. Layered Structures

One can make a superlattice consisting of alternating layers of CdTe and HgTe by use of molecular beam epitaxy. Such materials can be used as infrared detectors with properties that may be superior to $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ alloys (Berroir and Voos (1987)).

c. $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$

As this report reflects, $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ is the most important intrinsic semiconductor infrared detector material. $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$ has been considered as a possible competitor, because it may be more structurally stable. The use of $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$ is still quite new and much effort is being devoted to crystal growth techniques (Jozwikowski and Rogalski (1988)).

N89 - 21749

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

**MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA**

**SIMULATION OF WATER VAPOR CONDENSATION ON
LOX DROPLET SURFACE USING LIQUID NITROGEN**

Prepared by:	Eugene A. Powell
Academic Rank:	Senior Research Engineer
University and Department Engineering	Georgia Institute of Technology School of Aerospace
NASA/MSFC:	
Laboratory:	Propulsion
Division:	Propulsion Systems
Branch:	Performance Analysis
MSFC Colleague:	Richard Eskridge
Date:	September 2, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

SIMULATION OF WATER VAPOR CONDENSATION ON LOX DROPLET SURFACE USING LIQUID NITROGEN

by

Eugene A. Powell
Senior Research Engineer
School of Aerospace Engineering
Georgia Institute of Technology
Atlanta, Georgia

ABSTRACT

This project was concerned with the formation of ice or water layers on liquid oxygen (LOX) droplets in the Space Shuttle Main Engine (SSME) environment. Formation of such ice/water layers is indicated by phase-equilibrium considerations under conditions of high partial pressure of water vapor (steam) and low LOX droplet temperature prevailing in the SSME preburner or main chamber. An experimental investigation was begun using liquid nitrogen as a LOX simulant. A monodisperse liquid nitrogen droplet generator was developed which uses an acoustic driver to force the stream of liquid emerging from a capillary tube to break up into a stream of regularly spaced uniformly sized spherical droplets. The atmospheric pressure liquid nitrogen in the droplet generator reservoir was cooled below its boiling point to prevent two phase flow from occurring in the capillary tube. The cooling was accomplished by a jacket of liquid nitrogen boiling at subatmospheric pressure. An existing steam chamber was modified for injection of liquid nitrogen droplets into atmospheric pressure superheated steam. The droplets were imaged using a stroboscopic video system and a laser shadowgraph system.

Several tests were conducted in which liquid nitrogen droplets were injected into the steam chamber. Under conditions of periodic droplet formation, images of 600 micron diameter liquid nitrogen droplets falling through the steam were obtained with the stroboscopic video system. These pictures showed trails of submicron sized ice and/or water particles in the wakes of the liquid nitrogen droplets, but were unable to show ice or water layers on the surfaces of the particles. The laser shadowgraphs showed density gradients in the wakes of the particles, which were due to temperature and composition gradients caused by mixing between the cold nitrogen vapors given off by the droplets and the steam in the chamber. Recommendations are made for future studies utilizing improved diagnostic techniques and injecting the droplets into a high pressure steam environment.

INTRODUCTION

The life history of liquid oxygen (LOX) droplets is of critical importance to the performance of the Space Shuttle Main Engine (SSME). In the operation of liquid propellant rocket engines, the burning rate has been shown to be controlled by the vaporization of propellant droplets formed soon after injection [1]. This evaporation process normally includes a period in which the droplet is heated from its injection temperature to a steady-state wet-bulb temperature. The wet-bulb temperature is characterized by the transformation of all heat reaching the droplet surface into latent heat of vaporization with no further increase in temperature of the liquid. For droplets in a high pressure environment, such in a rocket thrust chamber, the wet-bulb temperature may not be reached. Under these conditions the droplet heating rate will control the droplet lifetime [2,3]. Combustion of LOX in the SSME is such an example. The droplet evaporation history determines the spatial distribution of energy release rate and also controls the combustion efficiency of the rocket.

The primary tool used to calculate the performance of the SSME is the ARICC code developed by Rocketdyne [4], which contains a submodel for droplet evaporation. In this submodel, it is assumed that the droplet heating process is quasi-steady and that phase-equilibrium prevails on the droplet surface. Heat is convected from the ambient gas to the droplet surface and carried away from the droplet surface due to mass transfer. The major driving force for LOX droplet evaporation is due to the difference between the O₂ vapor pressure at the liquid surface and the ambient O₂ partial pressure. The gas temperature and partial pressure distribution of O₂ vapor near the droplet surface are assumed to be continuous. Temperature ranges from the surface temperature (about 100 K) to ambient temperature (about 3000 K), and O₂ vapor pressure ranges from its saturated vapor pressure corresponding to the droplet temperature to the ambient O₂ vapor pressure.

In the SSME environment, the main combustion product is water vapor, which is about 75 percent by molar fraction while the balance is mainly hydrogen. Therefore the H₂O vapor pressure at the

LOX droplet surface will be very high, in the neighborhood of 2000 psi. This pressure is much higher than the saturated H_2O vapor pressure of ice at the liquid O_2 temperature of 100 K. With such a low temperature at the O_2 surface and very high H_2O vapor pressure, the phase-equilibrium concept implies that water vapor will condense and freeze at the liquid O_2 surface. Thus it is probable that there will be a layer of ice and liquid water surrounding the O_2 droplet during a part of the LOX droplet life history. Such an ice/water layer would invalidate the droplet evaporation submodel used in the ARICC code. The ice/water layer would be expected to extend the droplet lifetime thus reducing the evaporation rate and consequently the burning rate. This would have an adverse effect on the SSME performance.

On the other hand, if H_2O does not condense at the liquid O_2 surface, it implies that phase-equilibrium does not exist. In this case, super-saturated H_2O vapor will exist around the droplet surface, and the droplet evaporation submodel used in the ARICC code will not be adequate since it is based upon phase-equilibrium assumptions.

The possibility of the formation of ice/water layers on the LOX droplets in the SSME preburner and thrust chambers needs experimental verification. The work described in this report is concerned with the detection of ice and water layers on droplets of a cryogenic simulant injected into a chamber containing superheated steam at atmospheric pressure. This represents a first step toward the development of a high pressure facility to investigate the formation of condensed phase layers on cryogenic propellant droplets under SSME chamber conditions. The major part of the work described herein is concerned with the design, construction, and testing of a cryogenic liquid droplet generator to produce drops of uniform size and spacing.

OBJECTIVES

The objectives of this work were to:

1. Design and construct a monodisperse droplet generator for the production of liquid nitrogen droplets of uniform size and spacing.
2. Modify an existing steam chamber for use in the investigation of the formation of ice/water layers on liquid nitrogen droplets.
3. Explore possible diagnostic techniques for the detection of the ice/water layers on the liquid nitrogen droplets.

DEVELOPMENT OF THE MONODISPERSE LIQUID NITROGEN DROPLET GENERATOR

Although the formation of ice and water layers on liquid oxygen (LOX) droplets is of concern in the performance of the SSME, a cryogenic simulant was used in these studies to avoid the combustion and explosion hazards of liquid oxygen. The simulant should be inexpensive and have a droplet temperature similar to that of liquid oxygen. Argon has an atmospheric pressure boiling temperature of 87.5 K [5], which is very close to that of liquid oxygen (90.2 K), but liquid argon is too expensive. Liquid nitrogen was chosen because it is inexpensive, readily available and easily handled, and its atmospheric boiling temperature of 77.4 K is sufficiently close to that of liquid oxygen.

The requirements for the production of monodisperse liquid nitrogen droplets in a single stream of evenly spaced droplets arose from the desirability of using stroboscopic imaging techniques to "freeze" the droplets. Such techniques have been successfully used by other investigators in the study of burning fuel droplets. It is expected that the stroboscopic imaging method will also be useful in the investigation of the formation of ice and liquid water layers on the surfaces of cryogenic propellant droplets.

A prototype monodisperse droplet generator, which worked well with noncryogenic fluids such as water or ethanol, was previously developed in the Combustion Physics Laboratory by Richard Eskridge. This generator consisted basically of a cylindrical reservoir which fed liquid to a capillary tube. By applying an acoustic pressure disturbance to the surface of the liquid in the reservoir (using a simple loudspeaker), the liquid jet issuing from the capillary tube could be forced to break up into a stream of evenly spaced spherical droplets of uniform diameter. The droplets are produced at the rate of one for every cycle of oscillation of the acoustic pressure. In the absence of acoustic excitation, the stream still breaks up into droplets, but the drops are not uniform in size or regularly spaced. With acoustic driving the droplet formation process is periodic, without acoustic driving the droplet formation process is generally chaotic.

Attempts to produce liquid nitrogen droplets with this prototype droplet generator were unsuccessful due to two phase flow in the capillary tube. Due to excessive heat transfer from the environment, the liquid nitrogen in the capillary tube vaporized and caused the production of any droplets to be very intermittent and chaotic. Thus the droplet generator needed to be redesigned to minimize heat transfer to the liquid nitrogen both in the reservoir and the capillary tube.

A new droplet generator was designed and constructed in attempt to eliminate the two phase problem. A diagram of this apparatus and the associated acoustic driving and stroboscopic imaging systems are shown in Figure 1. The liquid nitrogen reservoir consists of a copper sphere about 20 cm in diameter with a capacity of approximately 4 liters. The sphere is insulated with a 25 mm thick layer of fiberglass wool and covered with aluminum tape.

ORIGINAL PAGE IS
OF POOR QUALITY

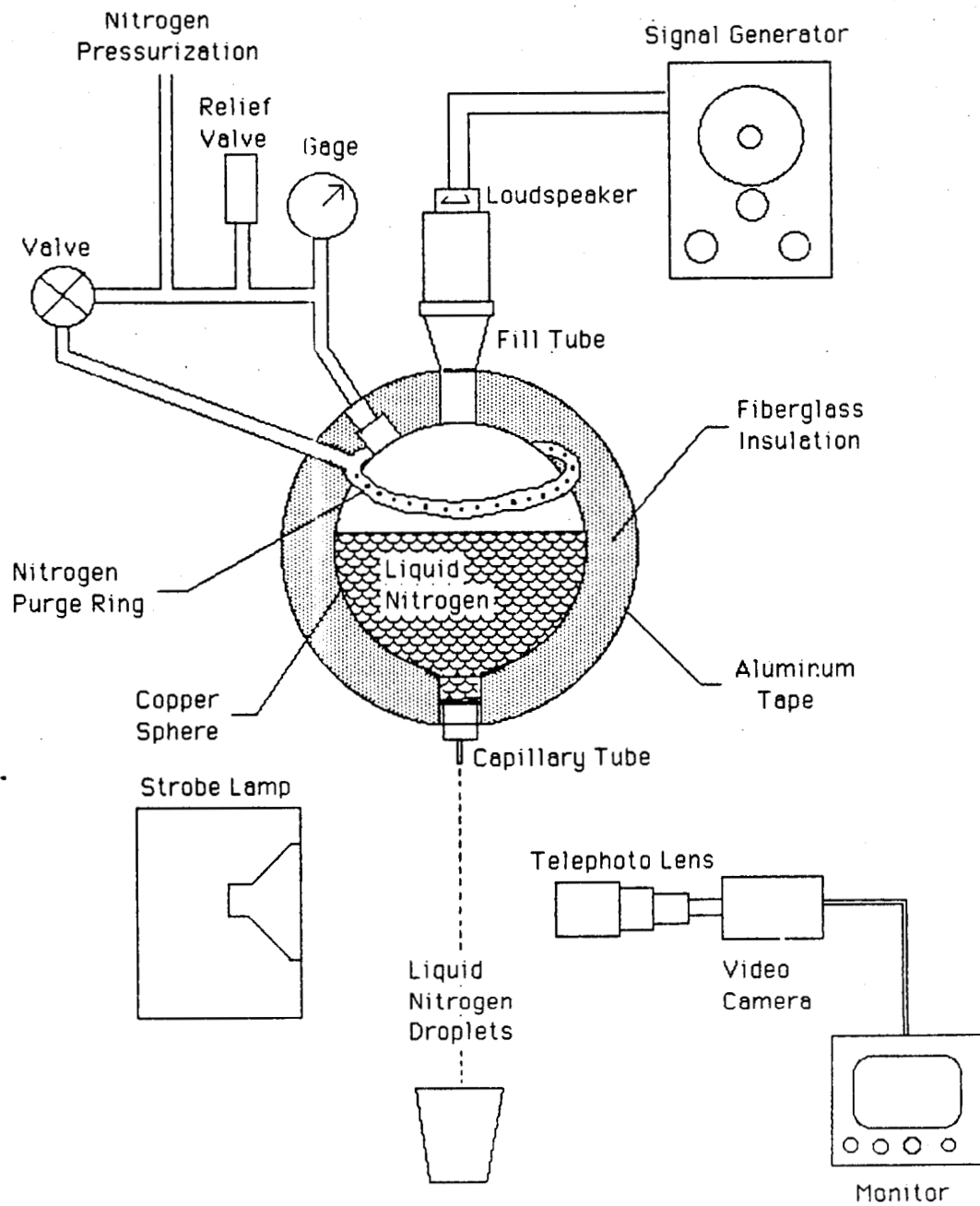


Figure 1. Liquid Nitrogen Droplet Generator, First Design

To prevent moisture from the air from condensing in the fiberglass insulation, a dry nitrogen purge is provided through a perforated ring of copper tubing embedded in the insulation. The nitrogen for this purge system is provided by boiloff vapor from the spherical reservoir. The reservoir is filled through a funnel-like galvanized iron pipe fitting at the top. A standard PVC pipe fitting containing the loudspeaker is screwed into the top of the galvanized fitting after the reservoir is filled. The droplet stream emerges from the bottom of the reservoir through either a stainless steel or glass capillary tube. The reservoir is provided with nitrogen gas pressurization in order to control the flow rate from the capillary tube.

The stroboscopic imaging system is also shown in Figure 1. The droplet stream is illuminated by a neon strobe lamp from behind, and the droplets are imaged using an extended telephoto lens on a video camera. The strobe light is synchronized with the framing rate of the video monitor. In order to obtain stroboscopic imaging of the droplets, the droplet generation frequency must be an integer multiple of the strobe frequency. For this system the loudspeaker output is sufficient to drive droplet formation for a frequency range of about 300 Hz to about 1200 Hz, while the video monitor framing rate is about 30 Hz. Therefore the droplet generation frequency ranges from about 10 to about 40 times the strobe frequency. Thus the acoustically driven droplet formation process must be very stable and the droplet stream must be protected from external disturbances such as air currents in order to obtain good stroboscopically frozen images.

Several tests of the droplet generator were conducted using water or ethanol as the fluid. In these tests acoustically driven periodic droplet formation and good stroboscopic imaging were obtained for several driving frequencies. Several operating characteristics were noted. By varying the nitrogen pressurization of the reservoir, the flow rate and hence the velocity of the fluid stream emerging from the capillary tube could be varied. Periodic droplet formation was sensitive to the velocity of the droplet stream. If the velocity was too low chaotic droplet formation prevailed. If the velocity was too high, the droplets were nonspherical and oscillated excessively. Under high velocity conditions small satellite droplets were also obtained. For a given acoustic driving frequency the spacing of the droplets increased as the liquid jet velocity increased.

Tests conducted with liquid nitrogen were unsuccessful in producing acoustically driven periodic droplet formation. The basic problem appeared to be boiling of the liquid nitrogen in the spherical reservoir and two phase flow in the capillary tube. Boiling in the reservoir may cause random disturbances to the droplet stream, and vapor bubbles in the reservoir are expected to damp the acoustic oscillations. Furthermore vapor bubbles entrained into the liquid flow through the capillary tube caused the liquid jet to be intermittent, a phenomenon which was easily seen using the imaging system. The results of these tests indicated that further measures are required in order to obtain acoustically driven periodic droplet formation with liquid nitrogen or other cryogenic fluids. In particular these measures must eliminate boiling in the liquid nitrogen reservoir and two phase flow in the capillary tube.

To prevent the liquid nitrogen in the droplet reservoir from boiling, it is necessary to cool it below its atmospheric boiling point of 77.4 K. This can be done by immersing the droplet reservoir in a low temperature bath obtained by boiling liquid nitrogen at subatmospheric pressure. The vapor pressure of liquid nitrogen as a function of temperature is given in Figure 2 [6]. In the absence of active refrigeration, cryogenic fluids such as liquid nitrogen always exist at their boiling point as a result of heat transfer from their surroundings. As Figure 2 shows, however, this boiling temperature can be lowered by reducing the pressure above the liquid. For example if liquid nitrogen is initially boiling at atmospheric pressure and the pressure is then reduced to approximately 0.5 atmosphere, the liquid will cool to a new boiling temperature of 72 K. During this cooling process an amount of liquid must be vaporized as determined by equating the latent heat of vaporization to the heat extracted from the remaining liquid during the cooling process. In principle a bath temperature as low as 64 K can be obtained by boiling liquid nitrogen at a pressure of about 0.15 atmosphere. For pressures below 0.12 atmosphere (the triple point of nitrogen), formation of solid nitrogen occurs.

Before designing a new monodisperse droplet generator using the subatmospheric nitrogen bath, a simple experiment was conducted to demonstrate the feasibility of this method. The apparatus for this experiment is shown in Figure 3. A 1000 ml Nalgene Erlenmeyer flask was nearly filled with liquid nitrogen, and a Pyrex glass bulb (50 ml capacity) was inserted into the liquid nitrogen bath. A side tube in the neck of the Erlenmeyer flask was connected by plastic tubing to a vacuum pump. About 40 ml of liquid nitrogen was then introduced into the glass bulb. One end of a length of plastic tubing was connected to the neck of the glass bulb, and the other end was submerged in a beaker of water. The boiling of the liquid nitrogen in the glass bulb could then be detected by the vapor bubbles emerging from the submerged tube. With the vacuum line disconnected, both the flask and the bulb were at atmospheric pressure and vigorous boiling occurred due to heat transfer from the environment.

When the vacuum pump was connected and turned on, the pressure in the flask dropped quickly to about 0.67 atmosphere, and boiling in the bulb ceased almost immediately. As the pressure in the flask gradually reached a steady state value of about 0.27 atmosphere, water was drawn up into the indicator tube. This was caused by nitrogen gas condensing in the glass bulb, thus reducing the pressure slightly below ambient pressure. This steady state condition was maintained for several minutes while the subatmospheric liquid nitrogen continued to boil. The temperature of the liquid nitrogen boiling at 0.27 atmosphere was about 67.7 K (Figure 2), which is 9.7 K below the atmospheric boiling point. At the end of the experiment the vacuum pump was shut off and the glass bulb was quickly removed and observed. The bulb was clear (no moisture was in the Erlenmeyer flask) allowing the liquid nitrogen in the bulb to be seen. The bulb still contained nearly 40 ml of liquid nitrogen which was not boiling. Shortly thereafter the bulb frosted over and vigorous boiling took place.

A new monodisperse liquid nitrogen droplet generator was designed and constructed using a cooling jacket of subatmospheric pressure liquid nitrogen.

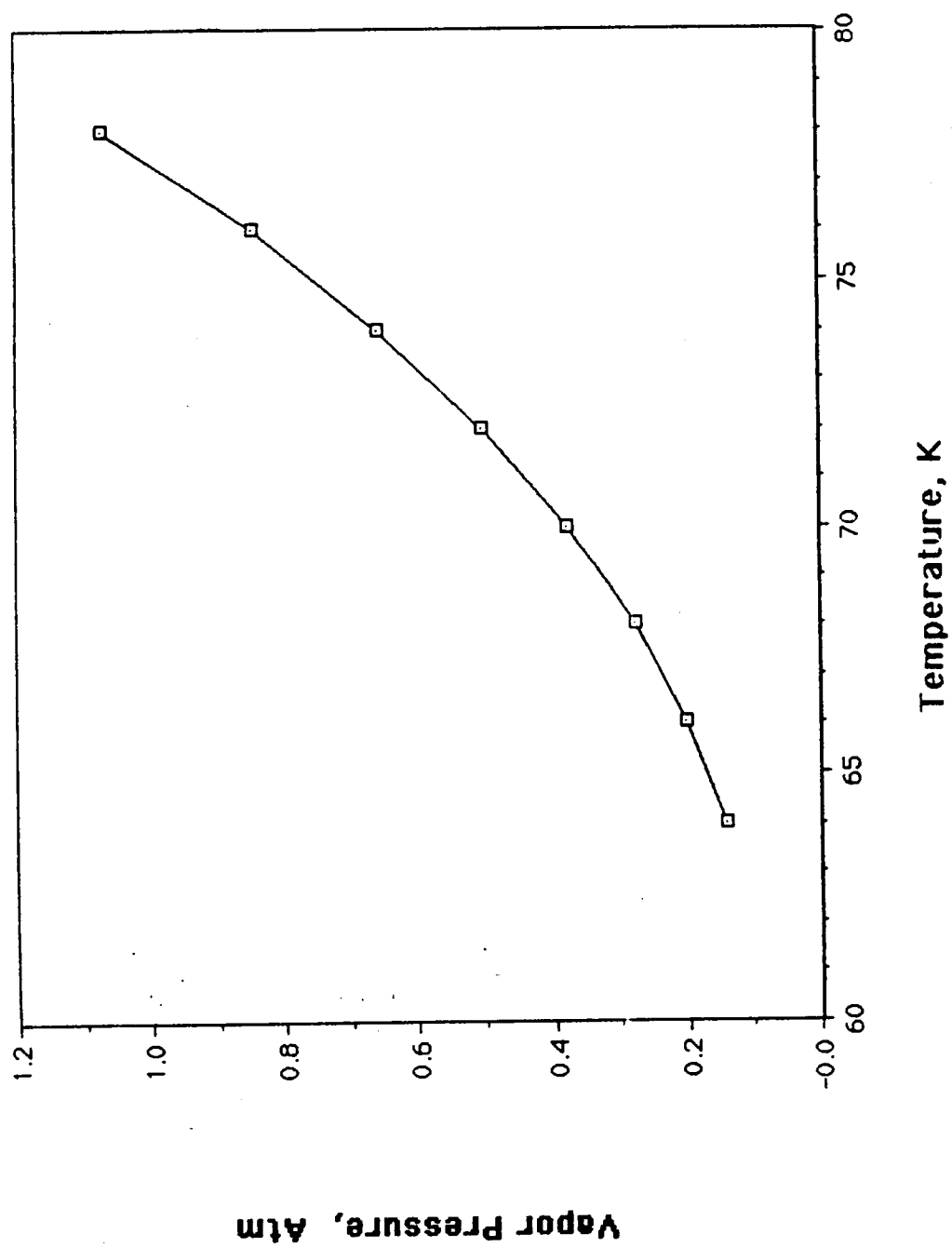


Figure 2. Vapor Pressure of Liquid Nitrogen vs Temperature

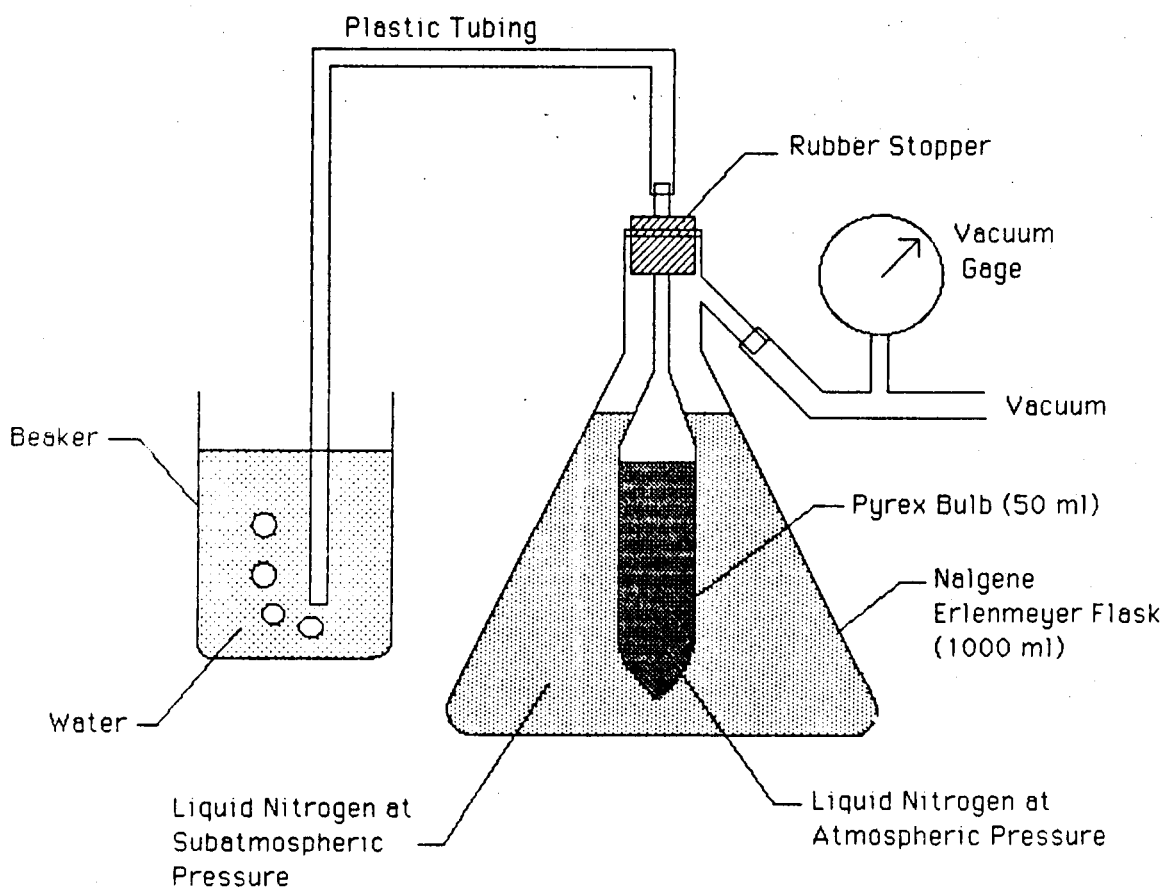


Figure 3. Demonstration Experiment for Cooling by Boiling Liquid Nitrogen at Subatmospheric Pressure

**ORIGINAL PAGE IS
OF POOR QUALITY**

A schematic diagram of this droplet generator is shown in Figure 4. The inner tube or droplet reservoir was constructed from a section of 3.8 cm diameter copper pipe about 33 cm long, while the outer jacket was made from 7.6 cm diameter copper pipe of approximately the same length. The capacities of the inner tube and the outer jacket are 350 ml and 1000 ml respectively. The entire outer jacket is insulated with a 2.5 cm thick layer of fiberglass wool covered with aluminum tape. The loudspeaker is connected to the upper end of the inner tube by a short length of PVC pipe. At the lower end of the inner tube a 2.5 cm length of 1.6 cm diameter copper pipe leads through the outer jacket to a threaded connection to which various capillary tubes can be attached.

The inner tube is provided with a venting and pressurization system through 0.63 cm diameter copper tubing. The venting is required during the cool down period when the liquid nitrogen in the inner tube is boiling. The vapors escape through the relief valve when the pressure exceeds about 2 psi. When the liquid nitrogen in the inner tube is subcooled and is not boiling, external pressurization from the laboratory gaseous nitrogen system is provided when necessary to obtain stable droplet formation. A water manometer is connected to this system to accurately measure this pressure, which ranges from zero to about 1.5 psi.

The outer jacket is provided with a vacuum system and a fill system. The vacuum line is connected to a vacuum gage and a valve for controlling the pressure in the cooling jacket. The vacuum is provided by two small vacuum pumps connected to a common manifold, giving a pumping capacity of about 400 liters/minute. The fill system consists of a reservoir connected by a valved 0.63 cm diameter insulated copper line. The insulated copper sphere from the earlier droplet generator was used for the jacket reservoir. The boiloff vapors from the spherical reservoir are used to provide a cold nitrogen purge for the insulation surrounding the droplet generator.

The following procedure was used for filling the droplet generator. With the jacket fill valve closed, the jacket reservoir is first filled using about 4 liters of liquid nitrogen from a small Dewar. The loudspeaker is then removed and the droplet reservoir is then filled. During the initial cooling down of the apparatus, considerable quantities of liquid nitrogen are boiled away, thus the droplet reservoir must be refilled several times. The valve is then opened to allow liquid nitrogen from the jacket reservoir to flow into the cooling jacket, and an additional 3 liters of liquid nitrogen is added to the jacket reservoir. The vacuum line is disconnected during this process, so that the jacket is at atmospheric pressure, and vapors from the liquid nitrogen boiling in the cooling jacket vent through the open vacuum line. When a spray of liquid nitrogen first emerges from the vacuum line, the cooling jacket is full and the flow from the jacket reservoir is shut off. Additional liquid nitrogen is then poured into the jacket reservoir and the droplet reservoir. A total of about 12 liters of liquid nitrogen is required to cool down and fill the droplet generator.

To operate the droplet generator, the vacuum line is then connected, and the vacuum pumps are started. After several minutes the jacket pressure stabilizes at about 0.67 atmosphere corresponding to a jacket temperature of

ORIGINAL PAGE IS
OF POOR QUALITY

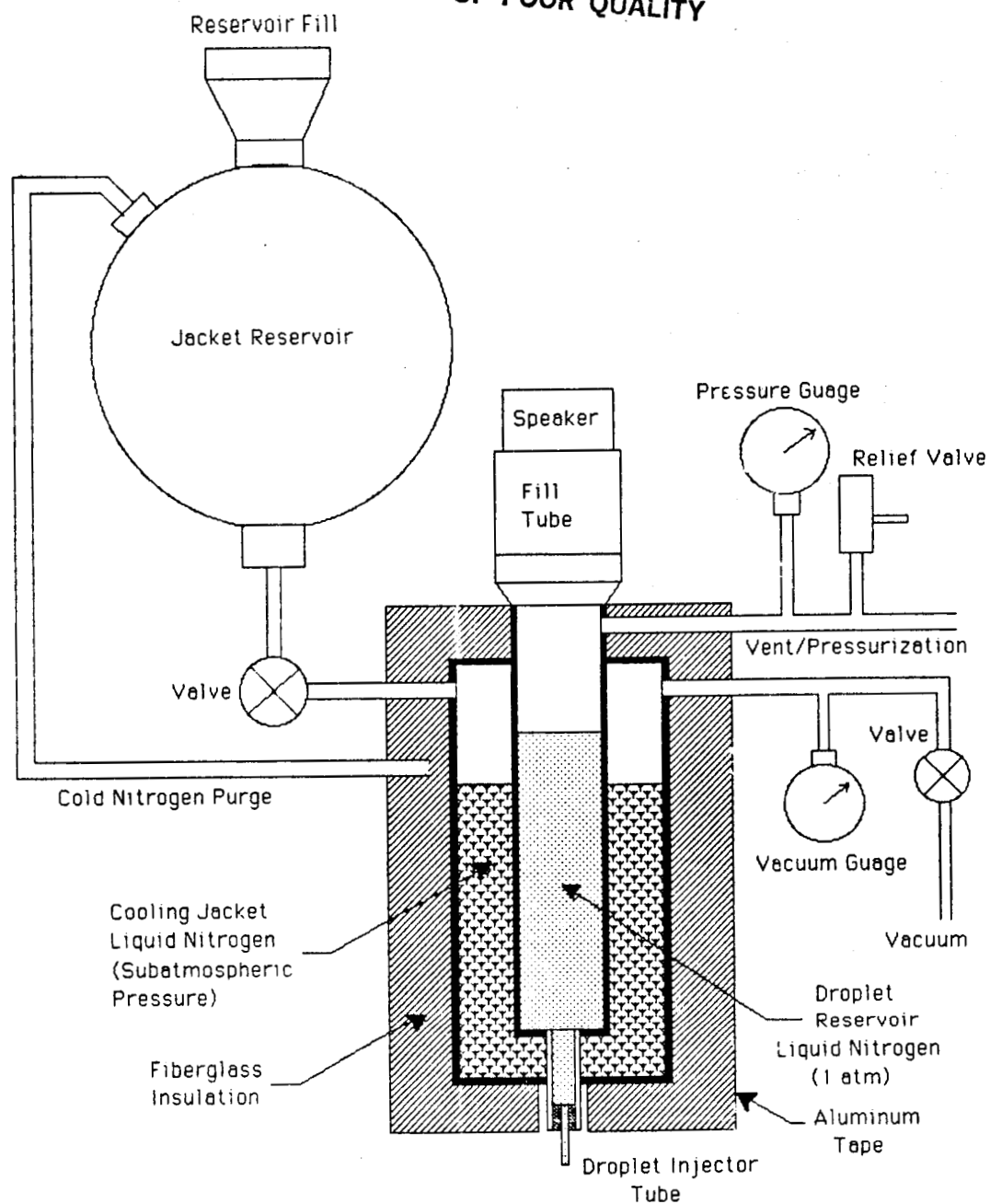


Figure 4. Monodisperse Liquid Droplet Generator, Second Design

about 74 K. At this point the droplet reservoir is again filled and the loudspeaker is replaced. As the liquid nitrogen in the inner reservoir is cooled the pressure in the droplet reservoir falls (measured with the manometer) as boiling ceases and some nitrogen recondenses. During steady state operation slight nitrogen pressurization of the inner reservoir is required to maintain the pressure about 0.1 psi above atmospheric pressure. This assures that a liquid nitrogen stream will always flow out of the capillary tube at the bottom of the droplet reservoir. During operation of the droplet generator, the valve from the jacket reservoir must be opened periodically to replace nitrogen boiling away in the cooling jacket.

The droplet generator was first tested using noncryogenic fluids such as water and ethanol. Acoustically driven periodic droplet formation was easily obtained over a frequency range of 300 Hz to 1300 Hz. Stroboscopically frozen images were obtained at driving frequencies which were integer multiples of the strobe frequency. Photographs of typical droplet streams are shown in Figure 5. The photograph on the left shows periodic droplet formation at about 600 Hz for water droplets. The droplets pinch off from the water stream at the top of the picture, where they are nearly spherical and uniform in size with a diameter of about 0.9 mm. As they fall they become noticeably ellipsoidal. The spacing between the drops is seen to be about 1.6 mm, which indicates a droplet velocity of about 1.0 m/sec. The photograph on the right shows water droplets produced without acoustic driving. These droplets are typical of chaotic droplet formation; they are produced in a range of sizes and at irregular intervals. The photographs in Figure 5 and others shown later in this report were obtained by playing a video recording of the droplet images into a digital image processing system. Using frame grabbing, individual video frames were displayed on the video monitor. These images were then photographed using Polaroid film in an oscilloscope camera.

Several test runs were made using liquid nitrogen in the droplet generator. In all of these tests the droplets were allowed to fall through room temperature air, and images were made of the drops a few centimeters below the capillary tube. During the most successful of these tests, monodisperse, acoustically driven periodic droplet formation was sustained for approximately one hour while various driving frequencies and droplet reservoir pressurizations were tried. With the droplet reservoir nearly full, the pressure head of the liquid nitrogen itself was usually sufficient for good droplet formation. At lower reservoir levels, pressurization with nitrogen or argon was needed; usually 0.5 to 1.5 psi was sufficient. The most uniformly sized droplets, which were also the most nearly spherical, were obtained with a driving frequency of 780 Hz (23 times the framing rate), but high quality droplets were also obtained at 720 Hz and 600 Hz. At this point in the test, the pressure in the cooling jacket was 0.77 atmosphere, corresponding to a boiling temperature of 75 K for the subatmospheric liquid nitrogen. This was apparently sufficient to eliminate boiling and two phase flow in the atmospheric pressure liquid nitrogen in the droplet reservoir, thus facilitating the production of monodisperse periodic droplets.

To further demonstrate that boiling of the liquid nitrogen in the droplet reservoir indeed prevents the formation of acoustically driven periodic

ORIGINAL PAGE IS
OF POOR QUALITY



Figure 5. Water Droplets Produced by the Monodisperse Liquid
Nitrogen Generator

droplets, the following procedure was carried out. The liquid nitrogen in the jacket was allowed to boil away with the jacket reservoir valve closed. Depletion of the liquid nitrogen in the jacket reservoir was indicated by a vacuum gage reading of about 0.05 atmosphere. At this point the droplet formation became chaotic; that is, it could not be synchronized with the strobe at any driving frequency. The jacket reservoir valve was then opened for about two minutes to allow liquid nitrogen to flow into the cooling jacket. Almost immediately acoustically driven periodic droplet formation was resumed. This was maintained for approximately four minutes while the liquid nitrogen in the jacket reservoir boiled away at about 0.5 atmosphere pressure (72 K). After the liquid nitrogen in the jacket reservoir was again exhausted, the droplet formation became increasingly random as the liquid nitrogen began to boil in the droplet reservoir.

MODIFICATION OF THE STEAM CHAMBER

A moderate pressure (80 psi) steam chamber used for a previous experiment in the Combustion Physics Laboratory at MSFC was modified for use as an atmospheric pressure steam chamber into which liquid nitrogen droplets could be injected. The modified steam chamber is shown in Figure 6.

The lower half of the existing chamber was retained essentially in its original form. This portion consists of two parts, a machined aluminum hexagonal observation section and a cylindrical lower chamber consisting of a flanged section of aluminum pipe. The observation section measures 35.6 cm across the flat faces and the central cylindrical cavity is 20.3 cm in diameter. The hexagonal block is fitted with six quartz windows 5.0 cm in diameter to provide optical access to the test chamber. The lower chamber also has an inside diameter of 20.3 cm and a length of 38.0 cm giving a total steam chamber height of 53.3 cm. The bottom of the lower steam chamber is closed with an aluminum plate with a 5 cm diameter central hole which allows steam to escape thus maintaining atmospheric pressure in the chamber. The lower chamber is wrapped with heating coils to maintain the steam chamber well above the atmospheric pressure boiling point of water (100 C) and thus prevent the condensation of liquid water in the chamber. The lower chamber and hexagonal block are insulated with about 2.5 cm thick layers of fiberglass wool wrapped in fiberglass tape. The hexagonal block contains no heating coils, but it is heated sufficiently by conduction from the lower chamber so that moisture does not condense on the optical windows. The steam is introduced into the lower chamber through a heated 1.3 cm diameter copper tube.

The upper portion of the steam chamber was modified for use with the monodisperse liquid nitrogen droplet generator. The top of the steam chamber is covered by a 28 cm diameter 3.2 cm thick aluminum plate, which is fitted with an O-ring seal and held down by six bolts. A 0.63 cm diameter hole at the center of the cover plate allows the liquid nitrogen droplets to enter the steam chamber. To prevent condensation of ice on the lower portion of the droplet generator and possible clogging of the capillary tube, a plenum chamber is used to separate the hot steam from the droplet generator. The plenum chamber consists of a flanged section of aluminum pipe which is bolted to the cover plate. The top of the plenum chamber is covered with a 0.3 cm thick circular aluminum plate with a 2.5 cm diameter central hole through which the insulated capillary tube is inserted. The plenum chamber provides a cylindrical cavity 10.2 cm in diameter and 7.3 cm deep which is purged with dry nitrogen to prevent steam from reaching the droplet capillary tube. The gaseous nitrogen enters the plenum chamber tangentially and disturbances to the droplet stream by the nitrogen flow are minimized by means of an internal baffle tube which extends to within about 1.0 cm of the top of the plenum chamber. The droplet generator fits tightly into the hole at the top of the plenum chamber so that all the nitrogen entering the plenum chamber exits through the hole at the bottom, thus preventing steam from entering the plenum chamber. The nitrogen flow rate is kept small, however, to prevent excessive dilution of the steam in the main chamber.

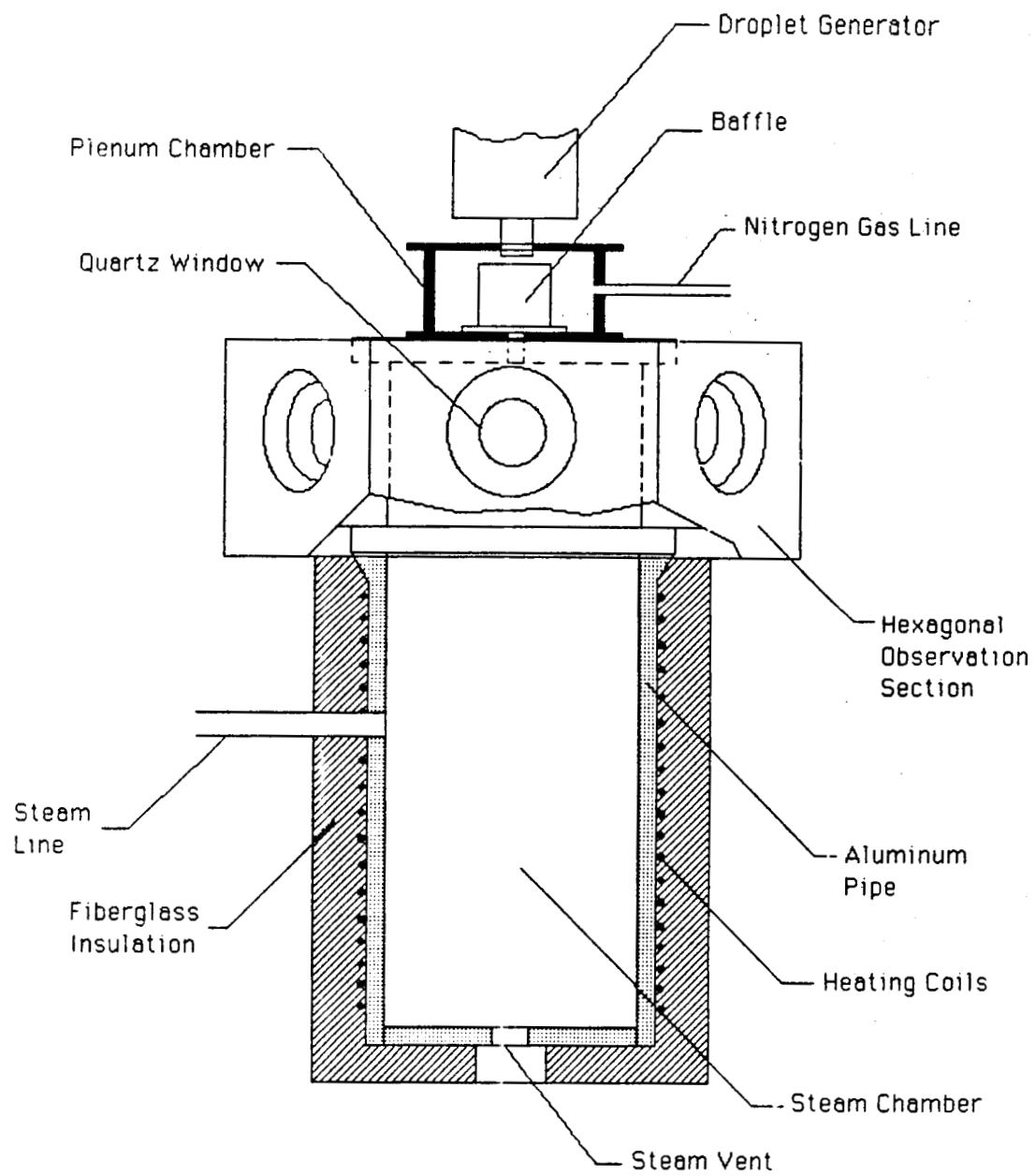


Figure 6. Atmospheric Pressure Steam Chamber

An existing boiler system was used to provide steam to the test chamber. This system is composed of a spherical stainless steel tank of approximately 38 liters internal volume wrapped with heating coils and heavily insulated. The boiler utilizes an internal pumping system to prevent thermal stratification in the boiler. The temperature of the boiler is controlled to an accuracy of 0.5 C using a platinum RTD thermocouple and controller. In the atmospheric pressure tests, the boiler temperature was held at about 98.4 C, and further heating of the steam occurred in the heated delivery line and the heated section of the steam chamber. During the steam tests, the test chamber surfaces were maintained at about 170 C using a separate thermocouple and controller.

RESULTS OF ATMOSPHERIC PRESSURE STEAM TESTS

Two tests were conducted in which liquid nitrogen droplets were injected into atmospheric pressure superheated steam using the monodisperse droplet generator and the modified steam chamber. The liquid nitrogen droplets were imaged using the stroboscopic video system used in previous tests of the droplet generator. In addition a stroboscopic laser shadowgraph system was used to view the droplets in this test.

The setup for the stroboscopic video system was slightly different than in previous tests due to the test cell geometrical constraints. The droplets could be viewed only at the fixed vertical location of the optical windows, which was about 18 cm below the droplet injector. Furthermore the camera could be placed no nearer than about 18 cm from the droplet stream, which was considerably farther than in previous tests in room temperature air. Consequently, the images of the droplets were considerably smaller than in the earlier tests.

The laser shadowgraph system was used to image density gradients in the vicinity of the droplets, which were invisible in the stroboscopic video images. This system utilized a 5 mW helium-neon laser (633 nm wavelength), which directed its beam to a rapidly rotating beam splitter cube. The cube was driven by a synchronous motor at 1800 rpm, giving a beam which scanned a small aperture at a frequency of 30 Hz, which is synchronous with the video framing rate. The duration of an individual pulse from the chopped beam was on the order of 5 μ sec, thus easily "freezing" the motion of the droplets. The pulsed beam then passed through beam expanding and collimating optics to produce a parallel beam of light passing through the observation section of the steam chamber. The pulsed beam then entered a second video camera to form an image of the shadows produced by droplets and density gradients in the test section.

In the atmospheric pressure steam tests it was difficult to obtain acoustically driven periodic droplet formation, and it was sustained only for a few minutes at a time at infrequent intervals. This difficulty may have been due in part to the increased heat transfer to the capillary tube in these tests, which may have resulted in two phase flow in the droplet injector. The top of the plenum chamber was hot to the touch, indicating that there was considerable heat transfer through the plenum chamber to the droplet generator in spite of the flow of cold nitrogen purge gas.

Typical stroboscopic video images of the liquid nitrogen droplets falling through atmospheric pressure steam are shown in Figure 7. In the photograph at the top, the droplets are being produced periodically by the acoustic driver; however, it appears that two parallel streams of droplets are being produced by the injector. This is an illusion produced by horizontal motions of the individual droplets under the influence of turbulence in the plenum chamber and the steam chamber. These flow disturbances made it virtually impossible to obtain clean stroboscopically frozen droplet chains as

ORIGINAL PAGE IS
OF POOR QUALITY

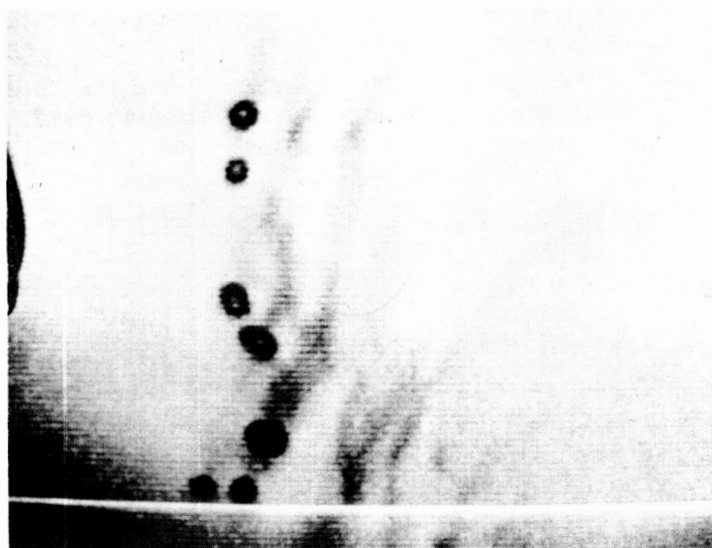
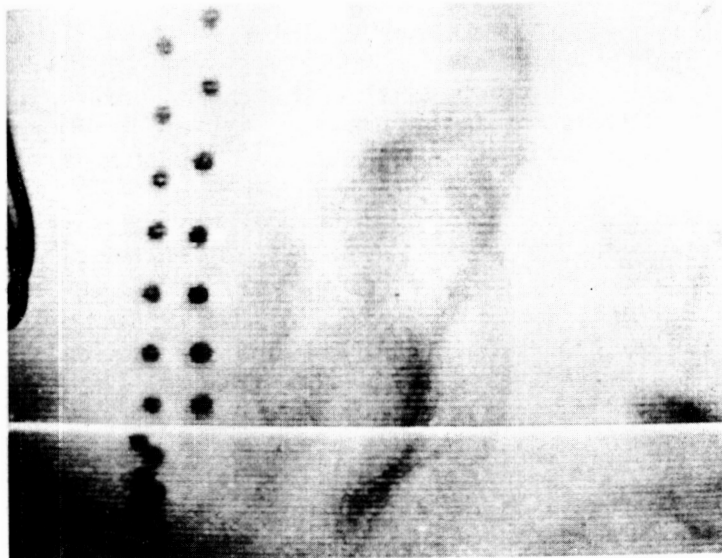


Figure 7. Stroboscopic Images of Liquid Nitrogen Droplets
Injected into Atmospheric Pressure Steam

were obtained in previous tests in atmospheric air (see Figure 5). The droplets in this photograph are approximately 600 microns in diameter. The lower photograph shows the droplets produced when the acoustic driver was unable to yield periodic monodisperse droplets. The droplets in the lower picture range in size from 600 to 1200 microns in diameter. The dark diffuse wisps to the right of the droplets in both photographs are due to the condensation of very small ice crystals and/or water droplets which occurs when the cold nitrogen vapors from the droplets mix with the hot steam in the chamber. These particles are produced predominantly in the wake of the falling droplets, forming comet-like tails several droplet diameters in length.

Images obtained with the laser shadowgraph system are shown in Figure 8. The upper photograph shows images of acoustically driven droplets produced synchronously with the laser pulse rate, while the lower photograph shows droplets obtained when periodic droplet production could not be maintained. Due to the small size of these droplets, the shadowgraph system does not produce sharp dark images of the droplets, but instead diffraction patterns are obtained. These patterns consist of concentric bright and dark rings which are dependent on the size and refractive index of the droplets. The oblique bands to the right of the droplets in the upper photograph are due to density gradients in the wakes of the three droplets shown and others below the field of view. These density variations are caused by temperature and composition gradients in the mixing regions in the wakes of the liquid nitrogen droplets. These density gradients are also visible in the wakes of the droplets shown in the lower photograph.

The images obtained of liquid nitrogen droplets injected into atmospheric pressure steam were unable to show any evidence of ice or water layers on the surfaces of the droplets. If any such layers of condensed phase water formed, they were probably only a few microns thick and were beyond the resolution of the imaging techniques employed. It is unlikely that the resolution of the direct imaging technique could be improved sufficiently to detect such thin ice/water layers. It is expected, however, that the diffraction patterns of the droplets may be affected by thin layers of ice and/or water sufficiently to be detected by the laser shadowgraph method if a higher power laser (5 W argon-ion) is employed and spatial filtering is used to clean up the background.

ORIGINAL PAGE IS
OF POOR QUALITY

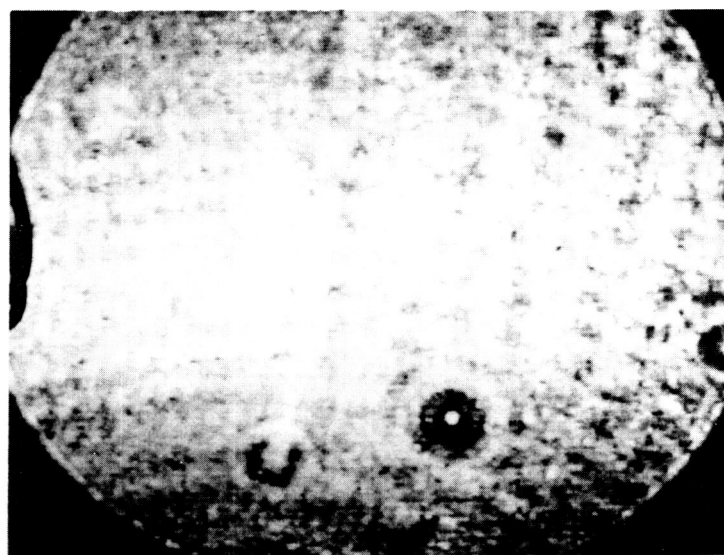
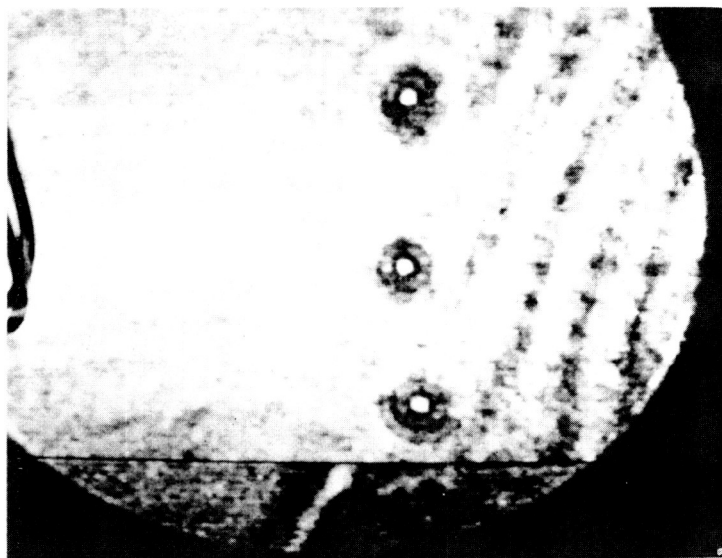


Figure 8. Laser Shadowgraph Images of Liquid Nitrogen
Droplets Injected into Atmospheric Pressure Steam

CONCLUSIONS AND RECOMMENDATIONS

In summary, most of the objectives of this project have been successfully met. A monodisperse liquid nitrogen droplet generator was constructed which produces liquid nitrogen droplets of uniform size at regular periodic intervals. This generator operates on the principle of acoustic excitation of an atmospheric pressure reservoir of liquid nitrogen cooled below its boiling point by means of a bath of liquid nitrogen boiling at subatmospheric pressure. The existing steam chamber was modified for the investigation of the formation of ice and water layers on liquid nitrogen droplets. Several tests in which liquid nitrogen droplets were injected into atmospheric pressure superheated steam were conducted. In these tests two diagnostic methods were evaluated: stroboscopic direct imaging of the droplets and pulsed laser shadowgraph imaging. Although neither diagnostic technique showed the presence of ice or water layers on the droplet surfaces, it is believed that such layers may be detectable by an improved version of the shadowgraph technique.

A new laser system has been recently acquired in the Combustion Physics Laboratory at MSFC which could be employed in the detection of ice and water layers on the liquid nitrogen droplets injected into steam. This system consisting of a tunable dye laser pumped by a pulsed xenon-chloride excimer laser has the potential of providing sufficient power to yield Raman imaging of the ice/water layers if they are present. The use of the stroboscopic imaging technique developed in this project would be advantageous here, since the Raman scattered intensity is expected to be small.

Based on the work conducted on this project the following recommendations are made:

The laser shadowgraph technique should be improved using a 5 W argon-ion laser and spatial filtering. Theoretical calculations of the effect of a thin ice or water layer on the diffraction patterns produced by the droplets are necessary for use of this diagnostic technique.

The pulsed dye laser system at MSFC should be used to obtain images of the ice and water layers by means of Raman scattering of the laser radiation. This system would be used in connection with the stroboscopic imaging technique for periodically produced monodisperse liquid nitrogen droplets.

A high pressure steam chamber should be designed and constructed which will provide better simulation of the SSME environment. This will require modifications of the liquid nitrogen droplet generator to enable the droplets to be injected at high pressure.

Other diagnostic techniques should be explored. For example measurements of scattered laser light at various angles coupled with Mie scattering theory may enable thin layers of water or ice to be detected on liquid nitrogen droplets, since the index of refraction of these materials are different. Theoretical investigations based on Mie scattering calculations for layered dielectric spheres are needed to determine the feasibility of this method.

REFERENCES

1. Priem, R. J. and Heidmann, M. F., "Propellant Vaporization as a Design Criterion for Rocket Engine Combustion Chambers", NASA TR R-67, 1960.
2. Wieber, P. R., "Calculated Temperature History of Vaporizing Droplets to the Critical Point," AIAA Journal, Vol. 1, No. 12, 1963, pp. 2764-2770.
3. Manrique, J. A. and Borman, G. L., "Calculations of Steady Droplet Vaporization at High Ambient Pressures", Journal of Heat and Mass Transfer, Vol. 12, 1969, pp. 1081-1095.
4. Liang, P. Y., Fisher, S., Chang, Y. M., "Comprehensive Modeling of a Liquid Rocket Combustion Chamber," AIAA paper 85-0232, AIAA 23rd Aerospace Sciences Meeting, Reno, Nevada, January 1985.
5. Handbook of Chemistry and Physics, 69th Ed., 1988-1989, CRC Press, Inc., Boca Raton, Fla.
6. Scott, R. B., Cryogenic Engineering, D. Van Nostrand Company, Inc., Princeton, N.J., 1959.

N89-21750

1988

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
UNIVERSITY OF ALABAMA

FURTHER INVESTIGATIONS OF
OBLIQUE HYPERVELOCITY IMPACT PHENOMENA

Prepared By:	William P. Schonberg
Academic Rank:	Assistant Professor
University and Department:	The University of Alabama in Huntsville Mechanical Engineering
NASA/MSFC:	
Laboratory:	Materials and Processes
Division:	Engineering Physics
Branch:	Laboratory Support
NASA Colleague:	Roy A. Taylor
Date:	August 8, 1988
Contract Number:	University of Alabama NGT-01-002-099

FURTHER INVESTIGATIONS OF
OBLIQUE HYPERVELOCITY IMPACT PHENOMENA

by

William P. Schonberg
Assistant Professor of Mechanical Engineering
University of Alabama in Huntsville
Huntsville, Alabama

ABSTRACT

This report describes the results of a continuing investigation of the phenomena associated with the oblique hypervelocity impact of spherical projectiles onto multi-sheet aluminum structures. A series of equations that quantitatively describes these phenomena is obtained through a regression of experimental data. These equations characterize observed ricochet and penetration damage phenomena in a multi-sheet structure as functions of the geometric parameters of the structure and the diameter, obliquity and velocity of the impacting projectile. Crater damage observed on the ricochet witness plates is used to determine the sizes and speeds of the ricochet debris particles that caused the damage. It is shown that, in general, the most damaging ricochet debris particle is approximately 0.25 cm (0.10 in.) in diameter and travels at speed of approximately 2.1 km/sec (6,890 ft/sec). The equations necessary for the design of shielding panels that will protect external systems from such ricochet debris damage are also developed. The dimensions of these shielding panels are shown to be strongly dependent on their inclination and on their circumferential distribution around the spacecraft. It is concluded that obliquity effects of high-speed impacts must be considered in the design of any structure exposed to the meteoroid and space debris environment.

INTRODUCTION

All spacecraft with a mission duration of more than a few days are susceptible to impacts by meteoroid and pieces of space debris. Such impacts are expected to occur at extremely high speeds and are expected to strike the spacecraft structure at oblique angles. High-speed oblique impacts are known to produce penetration and ricochet debris and can therefore damage internal as well as external flight-critical systems of the spacecraft. Either type of damage can in turn lead to catastrophic failure of the spacecraft and loss of life. Uncontained ricochet debris also increases the contamination of the orbital environment and can pose a threat to future missions into that environment. The design of a spacecraft for a long-duration mission must take into account the possibility of such impacts and their effects on the integrity of the entire structure. Protective systems for habitable portions and for external structural subsystems must be included in its design.

Protection against penetration for crew compartments and modules has traditionally consisted of a bumper plate that is placed at a small distance away from the main wall of the compartment or module. This concept was first proposed by Whipple [1] and has been studied extensively in the last few years as a means of reducing the penetration threat of hypervelocity impacts [2-15]. In the design process, bumper and wall plate thicknesses are iterated against weight and protection considerations to arrive at a final configuration.

In a recent investigation of hypervelocity impact [16], it was demonstrated that oblique high-speed impacts generate a tremendous volume of ricochet debris, especially for impact trajectories above 45 degrees. In fact, for trajectory obliquities beyond 60 degrees, the amount of penetration damage in a multi-sheet test specimen was observed to be minimal when compared to the damage sustained by the ricochet witness plate in the specimen. Unfortunately, previous investigations of oblique impact discuss only penetration phenomena and make little or no mention of damage induced by ricochet debris [2,9,15,17,18]. It has become clear that the damage potential of ricochet debris is very great, and that the creation of such debris is a dangerous phenomenon that deserves further attention.

The objectives of the research program performed under this fellowship were as follows: 1) to continue and extend the work begun in the first investigation by expanding the application of the equations previously developed; 2) to develop new equations for phenomena not discussed previously; 3) to analyze the formation and damage potential of ricochet debris and to develop a means of containing its spread. The results of this research program are presented in this report.

In the first section, the results of the previous investigation are reviewed. Suggestions for future research efforts made at its conclusion are also reviewed in light of the objectives of the present research program. In the next section, a review of the experimental procedure used in the oblique hypervelocity impact testing of multi-sheet specimens is presented. A complete set of impact test results is presented and reviewed. In the following sections, new equations governing the response of multi-sheet struc-

tures to oblique hypervelocity impact are presented. The damage potential of ricochet debris particles is analyzed by determining the sizes and speeds of typical ricochet debris particles. Equations and methodology for the design of shielding panels to protect external structural system elements are also developed. Several examples of panel design are presented. Finally, conclusions are made based on the data analysis of the preceding sections. Recommendations for future experimental and analytical investigations of hypervelocity impact are also presented.

RESULTS SUMMARY FROM PREVIOUS INVESTIGATION

Several important conclusions and recommendations were made at the completion of this author's first investigation of oblique hypervelocity impact phenomena [16]. These are reviewed below.

First, there exists a critical angle of obliquity. Projectiles with angles of obliquity less than this critical angle produce significant damage to the interior pressure wall and little damage to a ricochet witness plate. Projectiles with trajectory obliquities greater than the critical angle produce minimal damage to the pressure wall plate, but generate ricochet debris that causes major damage to a ricochet witness plate. This critical angle is estimated to have a value between 60 and 65 degrees. The existence of such an angle can have serious consequences on the design and placement of external subsystems such as instrumentation units on spacecraft that are developed for long-duration missions in the meteoroid and space debris environment.

Second, 99% of the ricochet debris particles generated traveled along trajectories never more than 30 degrees away from the bumper plate, regardless of the original angle of impact. The most serious ricochet damage was found to occur within an angle of 15 degrees with respect to the plane of the bumper plate, also regardless of impact angle. For original trajectory obliquities greater than 60 degrees, thin ricochet witness plates were completely perforated at the bumper plate/ricochet witness plate interface. In general, ricochet damage was found to increase with increases in original trajectory obliquity, original impact velocity, and the size of the original incident projectile.

Third, future experimental testing of oblique impact should be conducted with ricochet witness plates sufficiently thick so that little or no spalling or perforation occurs. With this stipulation, all the crater damage produced by ricochet particles can be used with thick plate equations to perform a detailed study of the damage potential of ricochet debris particles.

Fourth, future experimental investigations should also be conducted with larger diameter projectiles and specimen plates made from different thicknesses and materials. In this manner, the testing will better simulate the orbital space debris environment and the effect of bumper thickness on penetration and ricochet damage can be analyzed. Use of a wide variety of diameters and thicknesses will also serve to expand the applicability of current empirical expressions.

Fifth, more impact testing is needed at higher angles of obliquity to complement the large number of tests that have been performed at smaller angles (ie. less than 45 degrees). In light of the existence of a critical obliquity angle near 60 degrees, these tests are essential to be able to fully understand the oblique impact process.

The test program developed for the current investigation focused on the recommendations made in the last three paragraphs. Tests were performed using larger projectiles, thicker ricochet witness plates and various bumper plate thicknesses. The majority of the tests were performed at high angles of obliquity, although data from several previous normal impact tests were also included in some aspects of the analyses. The effect of the new data on existing equations will be addressed in subsequent sections.

EXPERIMENTAL PROCEDURE AND RESULTS

The oblique hypervelocity impact testing of multi-sheet specimens was done at the Space Debris Simulation Facility of the Materials and Processes Laboratory at the Marshall Space Flight Center. The facility consists of a light gas gun with a 12.7 mm (0.5 in.) launch tube capable of launching 2.5 -12.7 mm (0.1-0.5 in.) projectiles of mass 4-300 mg (0.009-0.661 lbs) at velocities of 2-8 km/sec (6,500-26,246 ft/sec). Projectile velocity measurements were accomplished via pulsed X-ray, laser diode detectors, and a Hall photographic station. This facility is fully described in Ref. 19. A drawing of the multi-sheet specimen set-up is shown in Figure 1.

In each test, a spherical projectile of diameter D and velocity V impacted a bumper plate of thickness t_b at an angle of obliquity θ . The projectile was shattered upon impact and created an elliptical hole in the bumper plate. Some secondary projectile and bumper plate fragments were sprayed upon the pressure wall plate a distance S away while some fragments ricocheted and struck the ricochet witness plate (thickness t_r).

The angles θ_1 , θ_2 , γ_1 , and γ_2 are 'perforation angles'. The angles θ_1 and θ_2 denote the trajectories of the centers of mass of bumper and 'in-line' penetration fragments, respectively; the angles γ_1 and γ_2 represent the spread of these fragments. The angles α_c and α_{99} are 'ricochet angles' and denote the trajectory of the center of mass of the ricochet fragments and the angle below which lie 99% of the ricochet fragments, respectively.

The formation and growth of penetration and ricochet debris clouds are clearly visible in Figures 2 through 5. These figures show the various stages of the oblique impact process beginning with a pre-event photograph (Figure 2). In Figure 3, a ricochet debris cloud consisting of projectile and plate fragments is the first to form. The penetration debris cloud is subsequently produced by further plate fragmentation. Its motion is initially directed along the outward normal of the reverse side of the plate (Figure 4). The impact event progresses and the remainder of the projectile fragments exit the rear of the plate. The penetration debris cloud then acquires an additional component of motion parallel to the rear surface of

the plate. The net result is that the penetration debris cloud begins to move in a direction similar to that of the original projectile (Figure 5). All photographs courtesy of JSC.

The configurations of the test specimens and the conditions of impact were chosen to simulate the conditions of space debris impact as closely as possible and still remain within the realm of experimental feasibility. Kessler [20] states that the average mass density for pieces of orbital space debris less than 10mm (0.3937 in.) in diameter is approximately 2.8 gm/cm³, which is approximately the same as the density of aluminum. Thus, the projectiles used were solid 1100 aluminum spheres with diameters ranging from 4.75 mm to 9.2 mm (0.187 in. to 0.375 in.).

The bumper, pressure wall, and ricochet witness plates were made of 6061-T6, 2219-T87, and 2219-T87 aluminum, respectively. The thicknesses of the bumper plates were varied from 0.8128 mm to 2.032 mm (0.032 in. to 0.080 in.); those of the ricochet witness plates were varied from 2.54 mm to 25.4 mm (0.1 in. to 1.0 in.). Pressure wall plate thickness was held constant at 3.175 mm (0.125 in.). The bumper and pressure wall plates were separated by a constant distance of 101.6 mm (4.0 in.). The obliquity of the impact was varied from 30 degrees to 75 degrees, while the impact velocities ranged from 5.0 to 8.0 km/sec (16,400 to 26,246 ft/sec). Data from several normal impacts were also used in the analysis of bumper plate holes.

A total of 30 oblique tests and 10 normal tests were used to study penetration and ricochet phenomena. The full database derived from these tests is presented in Tables 1 and 2. In Table 1, the angles Θ_1 and Θ_2 were obtained by estimating the locations of the centers of mass of the bumper plate fragments and 'in-line' projectile fragments on the pressure wall plate. The cone angles δ_1 and δ_2 were obtained by measuring the width of the damage cluster in the direction of the line of flight and by determining its location with respect to the edge of the pressure wall plate ahead of the cluster. The angle α was obtained by determining the vertical location of the center of mass^c of the ricochet debris based on the vertical distribution of the holes, craters, etc. formed by the debris. The angle α_{99} was determined based on the height below which lay 99% of the holes, craters, etc. formed by the ricochet debris. In Table 2, the minimum and maximum dimensions of the bumper plate hole in an oblique impact (D_{\min} and D_{\max}) as well as the hole diameters for normal impacts (D) were measured directly from the bumper plates.

A detailed qualitative description of the various processes involved in an oblique high-speed impact, including photographs of the various damage mechanisms, may be found in Refs. 15,16,21, and 22. Visual inspection of the new impact test specimens revealed a variety of phenomena that were basically consistent with these previous observations.

BUMPER PLATE HOLE ANALYSIS

Elastodynamic theory predicts that as a hypervelocity projectile strikes a plate, the projectile and the portion of the plate surrounding the impact site will break up into many fragments [23]. A portion of the frag-

ments will ricochet while the remainder will travel along the general direction of the original projectile towards the pressure wall plate. In order to estimate the damage potential of the ricochet and penetration fragments, it is necessary to know the total volume of debris generated by the impact. A good estimate of the bumper plate fragment volume can be obtained by multiplying the area of the hole formed during the impact by the thickness of the bumper plate. Inspection of the test specimens revealed the bumper plate hole to be elliptical with the elongation in the direction of the original projectile trajectory. The bumper plate hole area can be, therefore, approximated as the product of the maximum hole dimension and the minimum hole dimension.

Although the smaller hole dimension was previously observed to be independent of obliquity, inspection of the bumper holes in the new test specimens revealed an increasing dependence on obliquity, especially in the high obliquity specimens. Furthermore, the original equations for the maximum dimension of the bumper plate hole did not correlate well with experimental data for very large angles of obliquity. The objectives of the analyses in this task were to 1) modify the existing equation for D_{\min} by including an dependence on trajectory obliquity, and 2) improve the accuracy of the current equation for D_{\max} , especially in the high obliquity regime. In addition, it was also decided to extend the applicability of both equations to include normal, as well as oblique, high-speed impacts.

The new equations were obtained through a standard multiple linear regression of the hole dimension data with the following results:

$$D_{\min}/d = 2.825(V/C)^{1.043} \cos^{0.283} \Theta (t_s/d)^{0.782} + 1.01 \quad (1)$$

$$D_{\max}/d = 1.250(V/C)^{0.851} e^{1.064 \Theta} (t_s/d)^{0.672} + 1.40 \quad (2)$$

where C is the speed of sound in the bumper plate material, and Θ is in radians. The averages and standard deviations of the prediction errors of the regression model are presented in Table 3 (columns 1 and 2, respectively). A measure of the accuracy of the regression equations, the correlation coefficient, is presented for each equation in column 3. It can be seen from this Table and from Figure 6 that the equations are a fairly good fit to the hole dimension data. The relatively large spread of the prediction errors for eqn. (2) is due to an inherent physical uncertainty in the maximum hole dimension, especially in holes produced by high obliquity impacts. It has been shown previously that a high obliquity impact will tear, as well as perforate, the bumper plate in the direction of the line-of-flight of the projectile [16]. The effects of this tearing process on the maximum hole dimension varied dramatically, even between similar impact test shots (note the difference in D_{\max} for Tests No. 231C and 231D).

A comparison of Figure 6 with the corresponding figure in Ref. 16 shows that the new equations have a much wider range of applicability and possess a higher degree of accuracy than the previously developed equations. It is noted that these equations are valid only for projectiles and plates made of

the same material, for $0.0853 < t_s/d < 0.4278$, $5.0 < V < 8.0$ km/sec, and for trajectory obliquities between 0 and 95 degrees.

It is also interesting to note that the coefficients and exponents of eqn. (1) are very similar the corresponding constants in the equation obtained by Maiden, et.al. [7] for hole diameters in thin plates under normal high speed impact. This equation is given below.

$$D/d = 2.40(V/C)(t_s/d)^{0.666} + 0.90 \quad (3)$$

where D is the diameter of the (circular) hole in the bumper plate.

PENETRATION DEBRIS CENTER-OF-MASS TRAJECTORY ANALYSIS

As in the case of the hole dimension equations, the equations for θ_1 and θ_2 obtained previously were updated by including the data from the new test specimens in the regression database. Empirical expressions for θ_1 and θ_2 were obtained as functions of projectile diameter, impact velocity and trajectory obliquity with the following results:

$$\theta_1/\theta = 0.184(V/C)^{0.290} \cos^{1.372} \theta (t_s/d)^{-0.488}, \quad 45^\circ < \theta < 75^\circ \quad (4)$$

$$\theta_2/\theta = 0.490(V/C)^{-0.056} \cos^{0.909} \theta (t_s/d)^{-0.626}, \quad 30^\circ < \theta < 75^\circ \quad (5)$$

The averages and standard deviations of the prediction errors and the correlation coefficients for each equation are presented in Table 4. The spread of the prediction errors was found to be somewhat large for these equations. This is probably due to the fact that it was often difficult to determine the exact boundaries of the 'normal' and in-line' debris crater clusters and their centers of mass. The actual values of the penetration angles are therefore seen to be somewhat dependent on the person performing the analyses. However, eqns. (4,5) are a significant improvement over the corresponding equations developed in the previous study [16]. These equations are more compact, and for the most part have a higher correlation with experimental results than the equations previously developed.

A comparison of predicted and actual values of θ_1 and θ_2 is presented in Figure 7. It is seen that the 'in-line' trajectory angle, θ_2 , is not a single-valued function of trajectory obliquity, θ . In fact, θ_2 varies directly with θ up to a critical value, θ_{cr} , between 60 degrees and 65 degrees and then decreases with further increases in θ . This reversal at $\theta = \theta_{cr}$ corresponds to a change in the location of the most severe damage from the pressure wall plate for $\theta < \theta_{cr}$ to the ricochet witness plate for $\theta > \theta_{cr}$.

It is again noted that eqns. (4,5) are valid only for projectiles and plates of the same material, for $0.0853 < t_s/d < 0.4278$, and for $5.0 < V < 8.0$

km/sec.

PENETRATION DEBRIS CLOUD CONE ANGLES

In an effort to assess the extent of penetration damage as well as location, equations relating the spread of penetration debris were developed. These equations, together with the center of mass trajectory equations, could be used to assess whether or not the debris clouds formed as the result of an oblique impact would overlap and concentrate their energy or separate and distribute their energy upon the pressure wall plate.

Inspection of damage pressure wall plates revealed that for trajectory obliquities below 30 degrees and above 65 degrees there was significant overlapping of the projectile and bumper plate debris clouds. However, for intermediate obliquities, whether or not there was any separation of the debris clouds depended on the original impact parameters. It is interesting to note that in the case of low trajectory obliquity, the overlapping of the debris clouds concentrated the debris into a much smaller volume and thereby increased the damage potential of the penetrating debris particles. However, in the high obliquity regime, because so few penetration particles were created, the overlapping of the debris clouds did not contribute significantly to the damage caused by the debris particles.

The equations for the cone angles of the debris clouds were obtained using standard multiple linear regression techniques with the following results.

$$\theta_1/\theta = 0.417(V/C)^{0.228} \cos^{0.225} \theta (t_s/d)^{-0.491}, \quad 45^\circ < \theta < 75^\circ \quad (6)$$

$$\theta_2/\theta = 2.539(V/C)^{1.217} \cos^{2.972} \theta (t_s/d)^{0.296}, \quad 30^\circ < \theta < 65^\circ \quad (7)$$

It is noted that the regions of applicability of the cone angle equations are indicative of the overlapping phenomena for low and high trajectory obliquities. The averages and standard deviations of the prediction errors and the correlation coefficients for eqns. (6,7) are presented in Table 4 where it is seen that the equations are a fairly good fit to the cone angle data. Once again, these equations are valid for projectiles and plates of the same material, for $0.0853 < t_s/d < 0.4278$, and for $5.0 < V < 8.0$ km/sec.

RICOCET ANGLE ANALYSIS

Empirical expressions for α_c and α_{99} obtained previously were updated through a regression of the expanded database with the following results.

$$\alpha_c/\theta = 0.033(V/C)^{0.982} \sin^{-3.215} \theta (t_s/d)^{-0.531}, \quad 45^\circ < \theta < 75^\circ \quad (8)$$

$$\alpha_{99}/\theta = 0.194(V/C)^{0.390} \sin^{-1.874} \theta (t_s/d)^{-0.235}, \quad 45^\circ < \theta < 75^\circ \quad (9)$$

Average prediction errors, standard deviations, and correlation coefficients are presented in Table 4. Although the average prediction errors are quite small, the spread of the prediction errors is also somewhat large for these equations. This is probably due to error in the regression data itself which can be attributed to several factors. First, the ricochet witness plates were finite in height, and, as a result, some ricochet debris particles escaped detection. Second, ricochet debris holes and craters were frequently observed to cluster and overlap, especially for original trajectory obliquities greater than 60 degrees. In these cases it was difficult to determine the exact number of holes or craters on the ricochet witness plate.

Plots of predicted and actual ricochet angle values are presented in Figure 8. It can be seen from these curves that α_c has a weaker dependence on θ than α_{99} , which decreases significantly as θ increases. This feature is indicative of the fact that the majority of the ricochet debris particles travel along trajectories that are very close to the bumper plate, regardless of the original angle of impact.

It is also noted that these equations are valid only for projectiles and plates of the same material and for $0.0853 < t_s/d < 0.4278$ and $5.0 < V < 8.0$ km/sec.

RICOCHET PARTICLE SIZE AND VELOCITY ANALYSIS

The next task in the analysis of the oblique impact test specimens was to determine the sizes and velocities of typical ricochet debris particles based on the crater damage found on ricochet witness plates. It was proposed to use equations for penetration depth in thick plates together with measured crater depths to determine, in an inverse fashion, the diameter and velocity of the debris particles that caused the craters. Visual inspection of damaged ricochet witness plates revealed several interesting features that address the validity of this method.

First, the surface openings of ricochet witness plate craters formed by debris impacts were very nearly circular, which is indicative of normal or near-normal impact trajectories. This observation is confirmed by the data in Table 1 which indicates that 99% of the ricochet debris impact obliquities are less than 30 degrees, regardless of the original angle of impact. Second, in the tests where the ricochet witness plates were thicker than the standard 2.54 mm (0.1 in.) the reverse sides of the plates remained smooth and undamaged even though the front sides exhibited significant crater damage. In these cases, the post-impact appearance of the ricochet witness plate was identical to that of a 'thick plate' subjected to the same impact loading. Based on these observations, the proposal to use thick plate equations is justified provided that the reverse side of the ricochet witness plate in which the crater depths are measured is smooth and undamaged (ie. no spall or dimpling).

Examination of existing penetration depth equations revealed a strong coupling between particle size and velocity effects. That is, the same size crater can be produced by a small particle traveling at a high speed or by a larger particle traveling at a slower speed. In order to have a unique solution for the particle size and speed, a second set of equations describing another measurable crater quantity was needed. A search of existing literature on cratering phenomena in hypervelocity impact suggested crater volume to be such a quantity. Thus, a crater volume equation used in conjunction with an equation for penetration depth could be used to solve uniquely for particle size and speed. Since it is more facile to measure the surface diameter of an impact crater than it is to determine its exact volume, the crater volume equations were rewritten in terms of surface diameter. The analysis then proceeded as follows.

First, penetration depths and surface diameters of the three largest craters on ricochet witness plates with undamaged rear surfaces were measured. In this manner, the diameters and velocities subsequently calculated would represent upper bounds on ricochet debris sizes and speeds. Next, the equations for penetration depth and crater diameter were paired. Each pair of equations was then solved for particle diameter and velocity in terms of all other parameters, such as density, yield strength, wave speed, etc. Substitution of the appropriate values for these parameters in these equations yielded an estimate for the size and speed of the particle that produced a particular crater. This calculation was performed for each crater using 12 penetration depth equations and 6 crater diameter equations. These equations, some rewritten for consistency, are listed in the Appendix. Theoretically, this should have resulted in 72 estimates for the diameter and 72 estimates for the velocity of each crater producing projectile.

However, in the process of pairing the penetration depth and crater diameter equations, it became evident that not all equation pairs were compatible. Due to the exponential form of the equations, certain combinations of equations led to powers of zero for an unknown diameter or velocity. These particular equation pairs, therefore, could not be used to solve for the unknown quantities. This situation is analogous to finding the intersection of two parallel lines in Euclidean geometry.

Furthermore, even though an equation pair did produce a solution, the resultant particle size occasionally exceeded that of the crater diameter, sometimes by a factor of three or four. However, it was previously shown that the heated material surrounding a high speed impact crater relaxes as it cools after the impact event. This causes a reduction in crater diameter and depth of approximately 20 to 25% [24,25]. Therefore, it is indeed conceivable that a crater could have been produced by a particle whose diameter exceeded the size of the surface opening, but it is unlikely that the particle diameter could have exceeded the surface diameter by more than 25%. As a result, particle diameter values greater than 1.25 times the crater surface diameter were rejected. These two considerations reduced the number of calculated values from 72 to approximately 25. The averages of the acceptable diameters and velocities for each particle were then assumed to be valid estimates of its actual diameter and speed.

Measured crater depths and surface diameters for appropriate impact

tests are provided in Table 5. The resulting estimated particle diameters and velocities corresponding to these depths and diameters are presented in Table 6. The average values of these diameters and velocities (plus and minus a standard deviation) are shown in Table 7 as functions of original projectile diameter and impact velocity. Lastly, using the penetration depth and crater diameter equations, the estimated particle diameters and speeds are used to calculate the depth of penetration and surface diameter of a crater that would form if such a particle were to strike a ricochet witness plate of the same material. The percent differences between the actual crater depths and surface diameters in Table 5 and those calculated using the equations in the Appendix and the corresponding particle diameters and impact velocities are shown in Table 8. An examination of Tables reveals several interesting points.

First, from Tables 6 and 7, high obliquity impacts and impacts by large projectiles produce larger ricochet debris particles than impacts at low obliquities or impacts by small projectiles. This is a quantitative verification of the qualitative statement that the severity of the ricochet damage is directly related to the trajectory obliquity and size of the original projectile [16]. However, average ricochet debris particle velocity does not seem to vary significantly with respect to impact obliquity and speed and is approximately equal to 2.1 km/sec (6,890 ft/sec). The average diameter of the ricochet debris particles is calculated to be approximately 0.25 cm (0.1 in.).

Second, from the prediction errors in Table 8, particle diameters and speeds calculated using this technique will probably yield surface diameters fairly close to the actual values, but will over-estimate penetration depths by an average of approximately 20%. However, since the average penetration depth error is positive, the error is on the side of safety. The larger penetration depth prediction error due to the fact that the penetration depth equations used in this study often differ from each other by as much as 30% in their prediction of penetration depth. Predictions of crater diameter, however, more closely approximated the actual values due to the fact that the crater diameter equations generally yielded values that were within 5 to 10% of each other.

DESIGN OF EXTERNAL DEBRIS CONTAINMENT SHIELDS

Development of Shield Concept

It has been shown that an obliquely incident hypervelocity projectile will produce ricochet debris particles that can severely damage external flight critical systems of a spacecraft. Such particles can be as large as 0.25 cm (0.1 in.) and can travel as fast as 2.1 km/sec (6,890 ft/sec). It is evident that in the event of an on-orbit impact the ricochet debris that is produced must be contained in order to guarantee the safety of the mission and to avoid jeopardizing the safety of future missions into the same environment.

Although the concepts and procedures developed in the following sections are applicable to any type of spacecraft geometry and orientation, for

the purposes of this investigation it is assumed that the spacecraft for which containment shielding is to be designed is cylindrical and is oriented such that its velocity vector is parallel to its longitudinal axis. For the International Space Station, this implies that the modules are configured as shown in Figure 9 (illustration courtesy of Boeing Aerospace Corporation [15]). In this figure, the shaded areas are those which are most likely to be impacted by orbital debris.

A cross-section of a typical module is given in Figure 10, showing only the module wall, bumper plate and an array of containment shield elements. A V-shaped arrangement of two rectangular panels was chosen as the basic shield element configuration due to its ability to contain ricochet debris that would be produced by impacts with an equal probability of occurring from 'above' or from 'below' the orbital plane. In the assumed module orientation, shield elements would run parallel to the longitudinal axis of the module and would be evenly spaced along the circumference of the exposed portion of the module. It is interesting to note that in such a design, each panel of the shield element will serve as a bumper for the adjacent element and will trap any secondary 'penetration' debris that may form as a result of a ricochet debris particle impact.

In the design of such a containment shield system, it is assumed that the radius R of the spacecraft is a known quantity. The unknown quantities that need to be determined are the element panel length l_r , panel thickness t_p , panel inclination ϕ_e , and the element separation ϕ_s . Acceptable design values for these parameters are limited by the following phenomenological considerations.

1) The most dangerous ricochet debris particles are formed by impacts whose trajectory obliquities are greater than 60 degrees. Therefore, it is required that the angular separation of the containment shield elements be such that any projectile impacting the spacecraft between two adjacent shield elements have a trajectory obliquity of no more than 60 degrees (Design Condition #1).

2) Low obliquity impacts produce ricochet debris particles that can travel along trajectories of 30 degrees with respect to the plane of the bumper plate. These trajectory obliquities decrease as the obliquity of the parent projectile increases. Therefore, it is required that the length l_r of each containment shield element panel be sufficient to trap ricochet debris formed from an impact with a low trajectory obliquity, ie. less than 45 degrees (Design Condition #2).

The design procedure will basically consist of selecting a panel length, panel inclination, and an element separation and then ensuring that the chosen values satisfy Design Conditions #1 and #2. Panel thickness is calculated using a technique currently employed in the design of dual-wall structures.

Design Equations

Consider a projectile about to strike the external bumper wall of a cylindrical spacecraft with radius R (Figure 11). In the following para-

graphs, the terms 'downstream' and 'upstream' refer respectively to the shield elements in front of and behind an actual or projected impact site. Let ϕ_m be the angular separation between the site of impact and a 'downstream' shielding element, γ the obliquity of its trajectory with respect to the outward normal of the spacecraft hull, and let γ_{\max} be the maximum value of all such obliquities. The maximum obliquity occurs when the trajectory of an incoming projectile grazes an 'upstream' shield element and terminates at the base of an adjacent 'downstream' element. Thus, when $\gamma = \gamma_{\max}$, then $\phi_m = 0$.

According to Design Condition #1, the shield elements must be spaced so that all trajectory obliquities are less than 60 degrees. In Figure 12, a projectile traveling along a trajectory with an obliquity γ grazes the outermost point D of shield element panel DE and impacts at Point A; point C lies at the foot of the perpendicular drawn from point D to the extension of radius EF; point B is the foot of the perpendicular from point C to the extension of radius AB. Then angle CFB = $\phi_s - \phi_m$, angle CED = ϕ_e , and angle DAB = γ . Let angle ECA = α , angle CDA = ϵ , and angle CAB = γ_o ; then angle CAD = $\gamma_o - \gamma$. Lastly, let CA = z, DA = u, CE = v, CB = y, and BA = x.

In triangle FBC,

$$\frac{R + x}{R + l_r \cos \phi_e} = \cos(\phi_s - \phi_m)$$

from which

$$x = (R + l_r \cos \phi_e) \cos(\phi_s - \phi_m) - R \quad (10)$$

Also,

$$\frac{y}{R + l_r \cos \phi_e} = \sin(\phi_s - \phi_m)$$

from which

$$y = (R + l_r \cos \phi_e) \sin(\phi_s - \phi_m) \quad (11)$$

And,

$$\gamma_o = \tan^{-1}(y/x), \quad z = (x^2 + y^2)^{1/2} \quad (12), (13)$$

In triangle DCE,

$$w = l_r \sin \phi_e, \quad v = l_r \cos \phi_e \quad (14a, b)$$

Applying the Law of Cosines to triangle DCA,

$$w^2 = u^2 + z^2 - 2uz \cos(\gamma_o - \gamma) \quad (15)$$

Applying the Law of Sines to triangle CAD,

$$\frac{z}{\sin \epsilon} = \frac{u}{\sin(90 - \alpha)} \quad (16)$$

In triangle ACF,

$$\alpha + \phi_s - \phi_m + 180 - \gamma_o - 180$$

from which

$$\alpha - \gamma_o - (\phi_s - \phi_m) \quad (17)$$

In triangle CAD,

$$\epsilon + 90 - \alpha + \gamma_o - \gamma - 180$$

from which

$$\epsilon - 90 + \gamma - (\phi_s - \phi_m) \quad (18)$$

Substituting eqns. (17,18) into eqn. (16), noting that $\sin(90 \pm \beta) = \cos \beta$, and rearranging yields

$$u = gz, \quad (19)$$

where

$$g = \cos[\gamma_o - (\phi_s - \phi_m)] / \cos[\gamma - (\phi_s - \phi_m)]. \quad (20)$$

Substituting eqn. (19) into eqn. (15) and combining eqns. (14a) and (15) yields

$$w^2 = g^2 z^2 + z^2 - 2gz^2 \cos(\gamma_o - \gamma) - l_r^2 \sin^2 \phi_e \quad (21)$$

Substituting for z according to eqn. (13) and rewriting eqn. (21) as an equation in terms of the variable g yields

$$g^2 - 2g \cos(\gamma_o - \gamma) - l_r^2 \sin^2 \phi_e / (x^2 + y^2) + 1 = 0 \quad (22)$$

Substituting for x and y according to eqns. (10) and (11), respectively, yields the equation

$$g^2 - 2g \cos(\gamma_o - \gamma) + 1 - l_r^2 \sin^2 \phi_e / [R^2 + (R + l_r \cos \phi_e)^2 - 2R(R + l_r \cos \phi_e) \cos(\phi_s - \phi_m)] = 0 \quad (23)$$

where $g(\gamma, \gamma_o, \phi_s, \phi_m)$ is given by eqn. (20) and

$$\gamma_o = \tan^{-1} \left\{ (R + l_r \cos \phi_e) \sin(\phi_s - \phi_m) / [l_r \cos \phi_e - (R + l_r \cos \phi_e) [1 - \cos(\phi_s - \phi_m)]] \right\} \quad (24)$$

From Figure 11, $\gamma = \gamma_{\max}$ when $\phi_m = 0$. To calculate γ_{\max} , set $\phi_m = 0$ in eqn. (23) and solve for γ_{\max} . If $\gamma_{\max} < 60$ degrees for the chosen values of l_r

and ϕ_e , then Design Condition #1 is satisfied.

According to Design Condition #2, shield element panels must be long enough to contain ricochet debris formed by low obliquity impacts. Panel length varies directly with the distance from the impact site to the foot of the panel. Maximum panel length will be achieved by considering a low obliquity impact trajectory that grazes the outermost point of an 'upstream' element panel and places the impact site as far away as possible from the adjacent 'downstream' element.

In Figure 13, point A is the point of impact on the bumper of a projectile traveling along such a trajectory; BD is the 'downstream' shield element panel that must be long enough to contain the ricochet debris within angle BAD; point C lies at the intersection of line AD and the extension of radius BF; point G is a point on the line tangent to the circle at point A; point H is a point on the extension of radius AF. Then angle AFC = ϕ_m , angle CBD = ϕ_e , angle CAG = 30 degrees, angle CAG = 120 degrees, angle ACF = $60 - \phi_m$, AF = BF = R, and BD = l_r . Let the trajectory obliquity (angle EAH) and the ricochet cone angle DAG both be equal to 30 degrees, and let CB = u.

Applying the Law of Sines to triangle CAF yields

$$\frac{u + R}{\sin 120} = \frac{R}{\sin(60 - \phi_m)} \quad (25)$$

Solving for u yields

$$u = R[1 - (\cos \phi_m - 0.577 \sin \phi_m)] / (\cos \phi_m - 0.577 \sin \phi_m) \quad (26)$$

Applying the Law of Sines to triangle CBD yields

$$\frac{u}{\sin(120 + \phi_m - \phi_e)} = \frac{l_r}{\sin(60 - \phi_m)} \quad (27)$$

Solving for l_r yields

$$l_r = R[1 - f(\phi_m, 0)] / f(\phi_m, \phi_e) \quad (28)$$

where

$$f(x, y) = \cos(x - y) - 0.577 \sin(x - y) \quad (29)$$

Thus, if the value of l_r for the 'downstream' panel calculated using eqn. (28) is less than the assumed value of l_r , then the assumed panel length is sufficient and Design Condition #2 is satisfied.

Panel thicknesses can be calculated using a modified form of the following equation for rear-sheet thickness of a dual-wall structure [26].

$$t_2/d = 5.08V^{0.278}(t_1/d)^{-0.528}(h/d)^{-1.39} \quad (30)$$

where t_1 and t_2 are the thicknesses of the first and second sheets, respec-

tively, and h is the separation between the two sheets. Since the two panels are not uniformly spaced, an equivalent 'separation distance' was calculated as follows. Based on previous examinations of damaged ricochet witness plates, it was assumed that the majority of ricochet particle impacts will occur within a distance of $l_r/2$ away from the base of the shield element. Using the relationship between arc length, angle, and radius, the separation distance was approximated by the following equation.

$$h = (l_r/2)(2\phi_e) = l_r\phi_e \quad (31)$$

In order to equally protect against impacts from above and from below the orbital plane, the element design is assumed to be symmetric with respect to the outward normal of the spacecraft. Therefore, letting $t_1 = t_2 = t_p$, substituting for h in eqn. (30) according to eqn. (31), and solving for t_p yields the following expression for panel thickness.

$$t_p/d = 2.897V^{0.182}(l_r\phi_e/d)^{-0.910} \quad (32)$$

Design Procedure

The procedure to be used for the design of containment shielding panels is as follows.

Step No. 1: Input assumed values for ϕ_s , ϕ_e , l_r , and R . Based on previous experience, satisfactory initial values for these parameters are $\phi_s = 10$ degrees, $\phi_e = 5$ degrees, and $l_r = 0.1R$ to $0.2R$.

Step No. 2: Calculate γ_{\max} using eqn. (23) with $\phi_m = 0$. Is $\gamma_{\max} < 60$ degrees? If yes, proceed to Step No. 3; if no, choose a smaller ϕ_s or a larger l_r and repeat this step.

Step No. 3: Calculate ϕ_m for the assumed values of ϕ_s , ϕ_e , and l_r using eqn. (23) with $\gamma = 30$ degrees.

Step No. 4: Calculate the length l_r of the 'downstream' shielding element panel using eqn. (28). Is the assumed length l_r greater than the 'downstream' length l_r ? If yes, proceed to Step No. 5; if no, choose a larger l_r or a smaller ϕ_s and go to Step No. 2.

Step No. 5: The values of ϕ_s , ϕ_e , and l_r are acceptable. Calculate t_p using eqn. (32).

Several examples of panel design using this procedure and the accompanying equations are presented in the next section.

Examples

A matrix of acceptable design values for ϕ_s , ϕ_e , l_r , and t_p is given in Table 9. Panel length values were obtained for a spacecraft radius of 2.235 m (7ft. 4in.); thickness values were obtained using the average upper bound values for ricochet particle diameter and speed, namely, $d = 0.25$ cm and $V = 2.1$ km/sec.

It can be seen in Table 9 that a small change in ϕ_e produces only a minor change in l_r , whereas a similar small change in ϕ_s results in a major change in l_r . However, panel thickness is seen to be strongly dependent on the angle of inclination as well as the separation angle.

CONCLUSIONS AND RECOMMENDATIONS

This section consists of conclusions formulated during the course of this study together with a summary of those presented in the previous investigation [16]. These observations must be considered in the design of spacecraft meteoroid and space debris protection systems.

There exists a critical angle of obliquity. Projectiles with angles of obliquity less than this critical angle produce significant damage to the pressure wall plate and little damage to the ricochet witness plate in a hypervelocity impact test specimen. However, the damage produced on the pressure wall plate by projectiles with trajectory obliquities greater than the critical angle is minimal compared to the damage sustained by the ricochet witness plate. This critical angle is estimated to have a value between 60 and 65 degrees.

Low obliquity hypervelocity impacts are potentially more dangerous to a pressurized spacecraft module than normal impacts with otherwise identical parameters. In a low obliquity impact, the penetration debris cloud is more concentrated than in a normal impact. Even a small obliquity will concentrate the kinetic energies of the penetration debris particles which in turn can cause penetration of the pressure wall.

High obliquity impact penetration debris clouds have a lower damage potential than low obliquity impact debris clouds for two reasons. First, the quantity of penetration debris in high obliquity impacts is much lower than the quantity of penetration debris in low obliquity impacts. Second, the bumper plate debris and projectile debris clouds separate in a high obliquity impact which serves to dissipate the kinetic energy of the debris particles. Empirical equations that can be used to estimate the location and the extent of penetration damage were developed based on penetration angle data. These equations can also be used to determine whether or not the bumper plate and projectile debris clouds will overlap and combine their penetrating potentials or separate and dissipate their kinetic energies.

However, high obliquity impacts have a very high potential for damage to external spacecraft systems because of the large volume of ricochet debris particles that they produce. The most serious ricochet debris damage was found to occur within an angle of 15 degrees with respect to the plane of the bumper plate, regardless of the original angle of impact. For trajectory obliquities greater than 60 degrees, the ricochet witness plate was completely perforated at the bumper plate/ricochet witness plate boundary. A set of empirical equations that predict the farthest ricochet debris particle trajectory and the trajectory of the center of mass of the ricochet debris particle cloud were developed based on ricochet angle data.

An analysis of ricochet witness plate crater damage revealed that ricochet debris particles can be as large as 0.25 cm (0.1 in.) in diameter and can travel as fast as 2.1 km/sec (6,890 ft/sec). It is imperative that the issue of ricochet debris particle damage to external spacecraft systems be addressed in the development of spacecraft destined for the meteoroid or space debris environment.

A design concept for an external ricochet debris containment panel system was developed. Panel dimensions were found to be strongly dependent on the spatial distribution of containment shield elements around an exposed portion of a spacecraft, the inclination of the panels with respect to an outwardly pointing normal, and the orientation of the spacecraft in its orbital plane.

Additional analytical and experimental investigations of oblique hypervelocity impact phenomena are strongly recommended. Additional analytical investigations would achieve several goals. First, they would provide verification of the empirical equations developed in this study. Second, they would provide reliable means of predicting ricochet damage through accurate estimates of ricochet particle sizes and velocities. Third, they would yield damage criteria that would be applicable in a variety of impact situations. Future experimental investigations using projectiles and plates from different or composite materials would better simulate on-orbit impacts of space debris and would also serve to improve and extend the applicability of current empirical equations.

ACKNOWLEDGMENTS

The author wishes to acknowledge the support of the NASA/ASEE Summer Faculty Fellowship Program along with Michael Freeman, the University of Alabama Director, and Ernestine Cothran, the MSFC Program Co-Director. The author's gratitude is also extended to Roy Taylor, Chief of the Laboratory Support Branch of the Materials and Processes Laboratory for his support and guidance. The author also wishes to express his appreciation to Raymond Gause, Chief of the Engineering Physics Division, for his suggestions in the development of the ricochet debris containment shield system, and to Hubert Smith of the Laboratory Support Branch for conducting the impact testing which made this report possible.

REFERENCES

1. Whipple, E.L., "Meteorites and Space Travel", Astron. Journal, v. 52, 1947, p. 5.
2. Burch, G.T., Multi-plate Damage Study, AF-ATL-TR-67-116, Air Force Armament Library, Elgin Air Force Base, Florida, 1967.
3. D'Anna, P.J., "A Combined System Concept for Control of the Meteoroid Hazard to Space Vehicles", J. Spacecraft, v. 2, n. 1, 1965, p. 33.

4. Lundeberg, J.F., Lee, D.H., and Burch, G.T., "Impact Penetration of Manned Spacecraft", J. Spacecraft, v. 3, n. 2, 1966, p. 182.
5. Lundeberg, J.F., Stern, P.H., and Bristow, R.J., Meteoroid Protection for Spacecraft Structures, NASA CR 54201, Washington, D.C., 1965.
6. McMillan, A.R., Experimental Investigations of Simulated Meteoroid Damage to Various Spacecraft Structures, NASA CR 915, Washington, D.C., 1968.
7. Maiden, C.J., Gehring, J.W., and McMillan, A.R., Investigation of Fundamental Mechanism of Damage to Thin Targets by Hypervelocity Projectiles, GM-DRL-TR-63-225, General Motors Defense Research Laboratory, Santa Barbara, California, 1963.
8. Maiden, C.J., and McMillan, A.R., "An Investigation of the Protection Afforded a Spacecraft by a Thin Shield", AIAA Journal, v. 2, n. 11, 1964, p. 1992.
9. Merzhievskii, L., and Urushkin, V., "Oblique Collision of a High-Speed Particle with a Shield", Combust., Explos., and Shock Waves, v. 16, n. 5, 1981, p. 551 (Translation).
10. Nysmith, C.R., Penetration Resistance of Double Sheet Structures at Velocities to 8.8 km/sec, NASA TN D-4568, Washington, D.C., 1968.
11. Riney, T.D., and Halda, E.J., "Effectiveness of Meteoroid Bumpers Composed of Two Layers of Distinct Materials", AIAA Journal, v. 6, n. 2, 1968, p. 338.
12. Swift, H.F., Bamford, R., and Chen, R., "Designing Space Vehicle Shields for Meteoroid Protection: A New Analysis", Adv. Space Res., v. 2, n. 12, 1983, p. 219.
13. Wallace, R.R., Vinson, J.R., and Kornhauser, M., "Effects of Hypervelocity Particles on Shielded Structures", ARS Journal, 1962, p. 1231.
14. Wilkinson, J.P.D., "A Penetration Criterion for Double-Walled Structures Subject to Meteoroid Impact", AIAA Journal, v. 7, n. 10, 1968, p. 1937.
15. Coronado, A.R., Gibbins, M.N., Wright, M.A., and Stern, P.H., Space Station Integrated Wall Design and Penetration Damage Control, Report No. D180-30550-1, Contract NAS8-36426, Boeing Aerospace Company, Seattle, Washington, 1987.
16. Schonberg, W.P., Taylor, R.A., and Horn, R.A., An Analysis of Penetration and Ricochet Phenomena in Oblique Hypervelocity Impact, NASA TM-100319, Washington, D.C., 1988.
17. Johnson, W.E., "Oblique Impact Calculations Using a 3-D Eulerian Code", Proceedings of the AIAA Hypervelocity Impact Symposium, AIAA Paper No. 69-353, 1969.

18. Summers, J.L., Investigation of High Speed Impact: Regions of Impact and Impact at Oblique Angles, NASA TN D-94, Washington, D.C., 1959.
19. Taylor, R.A., "A Space Debris Simulation Facility for Spacecraft Materials Evaluation", SAMPE Quarterly, v. 18, n. 2, 1987, p. 28.
20. Kessler, D.J., Orbital Environment for Space Station, JSC20001
21. Schonberg, W.P., and Taylor, R.A., "Analysis of Oblique Hypervelocity Impact Phenomena", Proceedings of the 29th AIAA/ASME/ASCE/AHS SDM Conference, AIAA Paper No. 88-2370, 1988.
22. Schonberg, W.P., and Taylor, R.A., "Analysis of Oblique Hypervelocity Impact Phenomena for Meteoroid/Space Debris Protection System Design", Proceedings of the AIAA SDM Issues of the International Space Station Symposium, AIAA Paper No. 88-2463, 1988.
23. Cour-Palais, B.G., "Space Vehicle Meteoroid Shielding Design", Proceedings of the Comet Halley Micrometeoroid Hazard Workshop, ESA SP-153, 1979, p. 85.
24. Rolsten, R.F., and Hunt, H.H., "Impact Force and Crater Surface Area", AIAA J., v. 1, n. 8, 1963, p. 1893.
25. Rolsten, R.F., and Hunt, H.H., "Impact Force per Crater Area Related to Tensile Strength", J. Spacecraft, v. 1, n. 3, 1964, p. 351.
26. Nysmith, C.R., "An Experimental Impact Investigation of Aluminum Double-Sheet Structures", Proceedings of the AIAA Hypervelocity Impact Conference, AIAA Paper No. 69-375, 1969.
27. Sawle, D.R., "Hypervelocity Impact in Thin Sheets and Semi-Infinite Targets at 15 km/sec", AIAA J., v. 8, n. 7, 1970, p. 1240.
28. Bruce, E.P., "Review and Analysis of High Velocity Impact Data", Proceedings of the Fifth Hypervelocity Impact Symposium, 1962, p. 439.
29. , E., "Velocity Dependence of Some Impact Phenomena", Proceedings of the Comet Halley Micrometeoroid Hazard Workshop, ESA SP-153, 1979, p. 101.
30. Goodman, E.H., and Liles, C.D., "Particle-Solid Impact Phenomena", Proceedings of the Sixth Hypervelocity Impact Symposium, 1963, p. 543.
31. Dunn, W.P., "On Material Strengths of the Hypervelocity Impact Problem", AIAA Journal, v. 4, n. 3, 1966, p. 535.
32. Sedgwick, R.T., Hageman, L.J., Herrmann, R.G., and Waddell, J.L., "Numerical Investigations in Penetration Mechanics", Int. J. Engng. Sci., v. 16, 1978, p. 859.
33. Christman, D.R., "Target Strength and Hypervelocity Impact", AIAA Journal, v. 4, n. 10, 1966, p. 1872.

34. Summers, J.L., and Charters, A.C., "High-Speed Impact of Metal Projectiles in Targets of Various Materials", Proceedings of the Third Hypervelocity Impact Symposium, 1959, p. 101.

35. Sorenson, N.R., "Systematic Investigation of Crater Formation in Metals", Proceedings of the Fifth Hypervelocity Impact Symposium, 1962, p. 281.

36. Herrmann, W., and Jones, A.H., "Correlation of Hypervelocity Impact Data", Proceedings of the Fifth Hypervelocity Impact Symposium, 1962, p. 389.

37. Cour-Palais, B.G., "Hypervelocity Impact Investigations and Meteoroid Shielding Experience Related to Apollo and Skylab", Orbital Debris, NASA CP-2360, 1982, p. 247.

APPENDIX

Penetration Depth Equations

Reference No. 27:

$$p/d = 2.28(\rho_p/\rho_t)^{2/3}(V/C)^{2/3}, \quad V < 9 \text{ km/sec} \quad (P-1)$$

Reference No. 28:

$$p/d = 1.96(\rho_p/\rho_t)^{1/2}(V/C)^{2/3}, \quad V < 6 \text{ km/sec} \quad (P-2)$$

Reference No. 29:

$$p/d = 1.5(\rho_p/\rho_t)^{1/3}(\rho_p V^2/2s_t)^{1/3}, \quad V < 8 \text{ km/sec} \quad (P-3)$$

Reference No. 30:

$$p/d = 2.35(\rho_p/\rho_t)^{0.70}(V/C)^{2/3}, \quad V < 9 \text{ km/sec} \quad (P-4)$$

Reference No. 31:

$$p/d = 0.63(\rho_p V^2/\sigma_{yt})^{1/3}, \quad V < 7 \text{ km/sec} \quad (P-5)$$

Reference No. 32:

$$p/d = 0.482(\rho_p/\rho_t)^{0.537}(V/C)^{0.576}(Y_t/\rho_t C^2)^{-0.235}, \quad V < 21 \text{ km/sec} \quad (P-6)$$

Reference No. 33:

$$p/d = 8.355 \times 10^{-4} \rho_p^{2/3} \rho_t^{-1/3} (V^2/H_t)^{1/3}, \quad V < 9.5 \text{ km/sec} \quad (P-7)$$

Reference No. 34:

$$p/d = 2.0(\rho_p/\rho_t)^{4.52}(V/C)^{1.136}, \quad V < 9 \text{ km/sec} \quad (P-8)$$

Reference No. 35:

$$p/d = 0.311(\rho_p/\rho_t)^{0.17}(\rho_p v^2/s_t)^{0.285}, \quad V < 7 \text{ km/sec} \quad (\text{P-9})$$

Reference No. 36:

$$p/d = 0.36(\rho_p/\rho_t)^{2/3}(\rho_t v^2/B_t)^{1/3}, \quad V < 6 \text{ km/sec} \quad (\text{P-10})$$

Reference No. 37:

$$p = 2.973 \times 10^{-7} d^{1.1} H_t^{-0.25} \rho_p^{0.5} \rho_t^{-0.167} V^{4/3}, \quad V < \sim 3 \text{ km/sec} \quad (\text{P-11})$$

$$p = 1.129 \times 10^{-6} d^{1.056} H_t^{0.25} \rho_p^{0.5} \rho_t^{-0.167} E_t^{0.33} V^{4/3}, \quad V < \sim 3 \text{ km/sec} \quad (\text{P-12})$$

Crater Diameter Equations

Reference No. 18

$$\alpha d_h^2 p/d^3 = 34(\rho_p/\rho_t)^{3/2}(V/C)^2, \quad V < 4 \text{ km/sec} \quad (\text{C-1})$$

Reference No. 35:

$$\alpha d_h^2 p/d^3 = 0.120(\rho_p/\rho_t)^{0.5}(\rho_p v^2/s_t)^{0.845}, \quad V < 7 \text{ km/sec} \quad (\text{C-2})$$

Reference No. 28:

$$\alpha d_h^2 p/d^3 = 30.25(\rho_p/\rho_t)^{3/2}(V/C)^2, \quad V < 6 \text{ km/sec} \quad (\text{C-3})$$

Reference No. 30:

$$\alpha d_h^2 p/d^3 = 44.10(\rho_p/\rho_t)^{2/3}(V/C)^2, \quad V < 9 \text{ km/sec} \quad (\text{C-4})$$

Reference No. 33:

$$\alpha d_h^2 p/d^3 = 2.65 \times 10^{-9} \rho_p^{7/6} \rho_t^{-1/2} V^2/H_t, \quad V < 9.5 \text{ km/sec} \quad (\text{C-5})$$

Reference No. 36:

$$\alpha d_h^2 p/d^3 = 0.16(\rho_p/\rho_t)^{3/2} \rho_p v^2/B_t, \quad V < 6 \text{ km/sec} \quad (\text{C-6})$$

Notation

d_h crater surface diameter (cm)

d projectile diameter (cm)

p crater depth (cm)

B_t target material Brinell Hardness (dynes/cm²)

C speed of sound in target material (cm/sec)
 E_t target material elastic modulus (GPa)
 H_t target material Brinell Hardness Number (kg/mm²)
 S target material static shear strength (dynes/cm²)
 S_t target material dynamic hardness (dynes/cm²)
 Y_t target material dynamic shear strength (dynes/cm²)
 V projectile impact velocity
 α crater shape factor
 $\alpha = 0.75$ if $p > d_h/2$
 $\alpha = 1.00$ if $p \leq d_h/2$
 ρ_p projectile material mass density (gm/cm³)
 ρ_t target material mass density (gm/cm³)
 σ_{yt} ... target material dynamic yield strength (dynes/cm²)

Material Properties

$B_t = 1.27 \times 10^{10}$ dynes/cm ² $C = 5.10 \times 10^5$ cm/sec $E_t = 7.38 \times 10^{10}$ N/m ² $S = 2.83 \times 10^9$ dynes/cm ² $S_t = 6.37 \times 10^{10}$ dynes/cm ²	$Y_t = 2.78 \times 10^9$ dynes/cm ² $\rho_p = 2.71$ gm/cm ³ $\rho_t = 2.84$ gm/cm ³ $\sigma_{yt} = 1.85 \times 10^{10}$ dynes/cm ² $H_t = 130$ kg/mm ²
---	---

Test No.	V (km/sec)	d (mm)	Θ (deg)	Θ_1 (deg)	Θ_2 (deg)	δ_1 (deg)	δ_2 (deg)	α_c (deg)	α_{99} (deg)
EH1A	7.07	7.95	30	****	24.8	****	56.3	****	****
EH1B	6.96	7.95	45	10.9	38.1	42.3	30.7	15.5	29.2
EH1C	7.14	7.95	60	9.6	50.0	46.6	16.9	11.2	27.6
EH1D	7.18	7.95	75	4.7	26.9	57.1	****	7.9	27.1
EHAP	6.82	7.95	75	5.0	22.0	55.7	****	9.3	28.7
EHAA	6.93	7.95	75	4.7	22.2	47.8	****	9.5	30.1
EHAB	6.91	7.95	75	5.0	19.9	55.9	****	7.3	24.7
EHBP	7.22	6.35	75	4.3	21.8	48.7	****	6.6	26.0
EHCP	7.58	4.75	75	4.7	20.9	38.6	****	8.2	25.6
1061	6.84	8.89	60	11.3	47.1	43.4	26.5	****	****
106A	6.66	9.53	60	9.2	48.4	47.4	16.9	****	****
135C	6.76	6.35	30	****	24.0	****	53.0	****	****
135D	6.93	6.35	30	****	27.0	****	53.3	****	****
136A	6.25	6.35	55	10.7	43.5	43.9	20.9	8.7	23.3
136B	7.30	6.35	55	10.1	41.8	41.7	23.5	11.9	28.3
136C	6.67	6.35	55	11.0	38.2	44.5	24.3	12.9	28.4
150A	7.08	6.35	45	10.0	39.0	39.6	28.4	11.0	24.0
157A	7.40	4.75	60	9.3	36.0	42.4	21.2	8.0	22.0
162A	6.49	4.75	30	****	21.0	****	64.8	****	****
162B	5.03	4.75	30	****	27.0	****	52.7	****	****
206F	6.24	4.75	45	8.0	31.0	29.4	31.0	8.0	21.0
208E	6.48	6.35	65	9.0	47.0	43.5	11.7	8.0	20.0
209D	7.40	6.35	65	****	****	****	****	11.0	27.0
230C	5.16	6.35	45	10.0	34.0	34.3	23.7	11.0	26.0

230D	5.59	6.35	45	10.0	37.0	34.8	24.8	10.0	25.0
230E	6.62	6.35	45	10.0	32.0	33.0	28.3	12.0	25.0
231C	6.59	7.95	65	8.7	55.7	47.1	10.2	8.4	20.4
231D	7.26	7.95	65	10.2	49.7	48.5	20.1	9.7	23.0
EHRP1	6.87	7.95	60	10.6	46.5	52.9	22.6	10.6	23.6
EHRP2	6.80	7.95	65	11.0	64.4	49.3	9.4	8.7	19.1

Table 1 -- Impact Test Database: Penetration and Ricochet Angles

Test No.	V (km/sec)	d (mm)	θ (deg)	D _{min} (mm)	D _{max} (mm)	Eccen- tricity	D (mm)
EH1A	7.07	7.95	30	16.0	17.0	1.06	****
EH1B	6.96	7.95	45	16.5	20.0	1.22	****
EH1C	7.14	7.95	60	16.5	24.9	1.51	****
EH1D	7.18	7.95	75	14.5	36.1	2.49	****
EHAP	6.82	7.95	75	13.0	33.0	2.53	****
EHAA	6.93	7.95	75	13.2	33.5	2.54	****
EHAB	6.91	7.95	75	13.2	33.5	2.54	****
EHBP	7.22	6.35	75	10.9	23.1	2.09	****
EHCP	7.58	4.75	75	10.0	18.0	1.82	****
006A	6.95	6.35	0	****	****	1.00	15.0
013B	6.15	6.35	0	****	****	1.00	14.0
033	7.21	6.35	0	****	****	1.00	13.2
033C	5.53	6.35	0	****	****	1.00	11.2
035C	5.72	8.89	0	****	****	1.00	16.0
102	7.20	7.62	0	****	****	1.00	18.0
1061	6.84	8.89	60	18.8	29.0	1.54	****
106A	6.66	9.53	60	19.8	32.5	1.64	****
107B	6.82	8.89	0	****	****	1.00	18.5
135C	6.76	6.35	30	13.2	14.2	1.08	****
135D	6.93	6.35	30	13.2	14.2	1.08	****
136A	6.25	6.35	55	14.0	18.3	1.31	****
136B	7.30	6.35	55	14.0	20.1	1.44	****
136C	6.67	6.35	55	13.5	17.0	1.26	****
150A	7.08	6.35	45	14.2	18.0	1.26	****

157A	7.40	4.75	60	13.7	17.3	1.26	****
162A	6.49	4.75	30	11.9	14.0	1.18	****
162B	5.03	4.75	30	9.9	11.7	1.17	****
206F	6.24	4.75	45	11.7	13.5	1.16	****
208E	6.48	6.35	65	13.0	21.0	1.61	****
209D	7.40	6.35	65	14.5	19.6	1.36	****
213B	5.90	7.95	0	****	****	1.00	16.5
228B	6.75	7.95	0	****	****	1.00	12.7
228D	6.65	6.35	0	****	****	1.00	11.2
230C	5.16	6.35	45	12.4	16.0	1.28	****
230D	5.59	6.35	45	13.5	16.3	1.22	****
230E	6.62	6.35	45	14.2	17.5	1.25	****
231C	6.59	7.95	65	16.5	31.0	1.87	****
231D	7.26	7.95	65	16.5	25.9	1.57	****
EHRP1	6.87	7.95	60	16.5	29.0	1.74	****
EHRP2	6.80	7.95	65	16.0	33.0	2.19	****

Table 2 -- Impact Test Database: Bumper Plate Hole Dimensions

	$\% \epsilon_{avg}$	$\sigma(\%)$	$100R^2$
D_{min}/d	0.001	6.550	72.3
D_{max}/d	0.055	11.410	86.1

Table 3 -- Regression Analysis of Bumper Hole
Dimension Data, Error Summary

	$\% \epsilon_{avg}$	$\sigma(\%)$	$100R^2$
θ_1/θ	0.612	11.029	94.8
θ_2/θ	2.209	21.436	73.9
δ_1/θ	0.187	6.261	73.8
δ_2/θ	2.515	22.436	91.5
α_c/θ	0.889	13.676	85.5
α_{99}/θ	0.640	11.832	75.6

Table 4 -- Angle Data Regression Analysis, Error Summary

Impact Parameters				Penetration Depths and Diameters					
Test Number	Velocity km/sec	θ deg	d cm	p_1 cm	d_{h1} cm	p_2 cm	d_{h2} cm	p_3 cm	d_{h3} cm
EHAB	6.91	75	0.795	0.615	0.734	0.368	0.636	0.483	0.566
EHBP	7.22	75	0.635	0.495	0.650	0.361	0.602	0.310	0.445
EHCP	7.58	75	0.475	0.386	0.599	0.318	0.447	0.345	0.422
EHRP2	6.80	65	0.795	0.371	0.632	0.229	0.445	0.211	0.445
EHRP5	7.51	65	0.635	0.305	0.528	0.330	0.546	0.203	0.411
EHRP1	6.87	60	0.795	0.140	0.254	0.094	0.241	0.117	0.244
EHRP4	7.65	60	0.635	0.152	0.279	0.216	0.371	0.157	0.328
EHRP7	7.98	60	0.475	0.323	0.488	0.254	0.396	0.203	0.465
EHRP3	6.78	45	0.795	0.165	0.368	0.150	0.320	0.135	0.343
EHRP6	7.57	45	0.635	0.097	0.201	0.114	0.267	0.084	0.211
EHRP8	7.34	45	0.475	0.155	0.262	0.137	0.279	0.168	0.295

Table 5 -- Measured Penetration Depths and Crater Diameters

Impact Parameters				Particle Diameters and Velocities					
Test Number	Velocity km/sec	Θ deg	d cm	d_1 cm	V_1 km/sec	d_2 cm	V_2 km/sec	d_3 cm	V_3 km/sec
EHAB	6.91	75	0.795	0.472	2.40	0.452	1.97	0.346	2.45
EHBP	7.22	75	0.635	0.389	2.58	0.395	2.06	0.254	2.50
EHCP	7.58	75	0.475	0.390	2.20	0.253	2.55	0.261	2.47
EHRP2	6.80	65	0.795	0.350	1.96	0.361	2.03	0.243	1.99
EHRP5	7.51	65	0.635	0.302	2.28	0.253	2.25	0.289	1.80
EHRP1	6.87	60	0.795	0.413	1.98	0.289	1.85	0.260	2.21
EHRP4	7.65	60	0.635	0.176	2.08	0.240	2.02	0.208	2.01
EHRP7	7.98	60	0.475	0.125	2.06	0.157	1.97	0.123	1.95
EHRP3	6.78	45	0.795	0.160	2.09	0.140	1.93	0.148	2.19
EHRP6	7.57	45	0.635	0.224	1.98	0.183	2.24	0.191	1.97
EHRP8	7.34	45	0.475	0.169	2.07	0.168	2.13	0.186	2.09

Table 6 -- Calculated Particle Diameters and Velocities

		d_{avg} (cm)	V_{avg} (km/sec)
(a)	45°	0.174 ± 0.024	2.07 ± 0.10
	60°	0.221 ± 0.087	2.01 ± 0.09
	65°	0.299 ± 0.044	2.05 ± 0.17
	75°	0.357 ± 0.079	2.35 ± 0.21
(b)	0.475	0.203 ± 0.080	2.17 ± 0.20
	0.635	0.258 ± 0.070	2.15 ± 0.19
	0.795	0.303 ± 0.110	2.08 ± 0.18

Table 7 -- Average Ricochet Particle Diameters and Velocities
 (a) as a Function of Original Projectile Obliquity
 (b) as a Function of Original Projectile Diameter (cm)
 (Original Projectile Impact Velocity ~ 7.3 km/sec)

Impact Parameters				Prediction Errors (%)					
Test Number	Velocity km/sec	Θ deg	d cm	p ₁	d _{h1}	p ₂	d _{h2}	p ₃	d _{h3}
EHAB	6.91	75	0.795	-6.0	16.7	28.0	5.4	-10.9	12.7
EHBP	7.22	75	0.635	1.8	13.8	17.9	7.8	3.0	6.8
EHCP	7.58	75	0.475	15.1	11.9	1.7	7.1	-6.2	14.6
EHRP2	6.80	65	0.795	19.0	6.2	16.3	7.7	24.2	-13.7
EHRP5	7.51	65	0.635	9.3	9.0	14.5	11.3	43.3	-14.8
EHRP1	6.87	60	0.795	16.5	4.8	24.1	-0.4	39.5	-8.7
EHRP4	7.65	60	0.635	24.1	4.5	17.4	5.4	38.5	-6.4
EHRP7	7.98	60	0.475	37.0	-7.0	40.2	-14.2	47.7	-16.1
EHRP3	6.78	45	0.795	23.1	4.6	50.1	-16.6	41.2	-5.6
EHRP6	7.57	45	0.635	40.2	-11.5	39.1	-9.8	45.4	-19.2
EHRP8	7.34	45	0.475	16.5	6.6	33.6	-8.3	19.6	4.9

Table 8 -- Prediction Errors Using Calculated Particle Diameters and Velocities

Panel Inclination (ϕ_e)								
5°			10°		15°		20°	
ϕ_s	l_r	t_p	l_r	t_p	l_r	t_p	l_r	t_p
5°	12.19	0.222	11.58	0.124	10.97	0.090	10.67	0.071
10°	27.43	0.106	25.91	0.060	24.38	0.044	23.47	0.035
15°	48.77	0.063	44.20	0.037	42.67	0.026	38.10	0.023
20°	73.15	0.043	67.06	0.025	60.96	0.019	56.39	0.016

Table 9: Panel Length (cm) and Thickness (cm) as a Function of Panel Inclination and Shield Element Separation

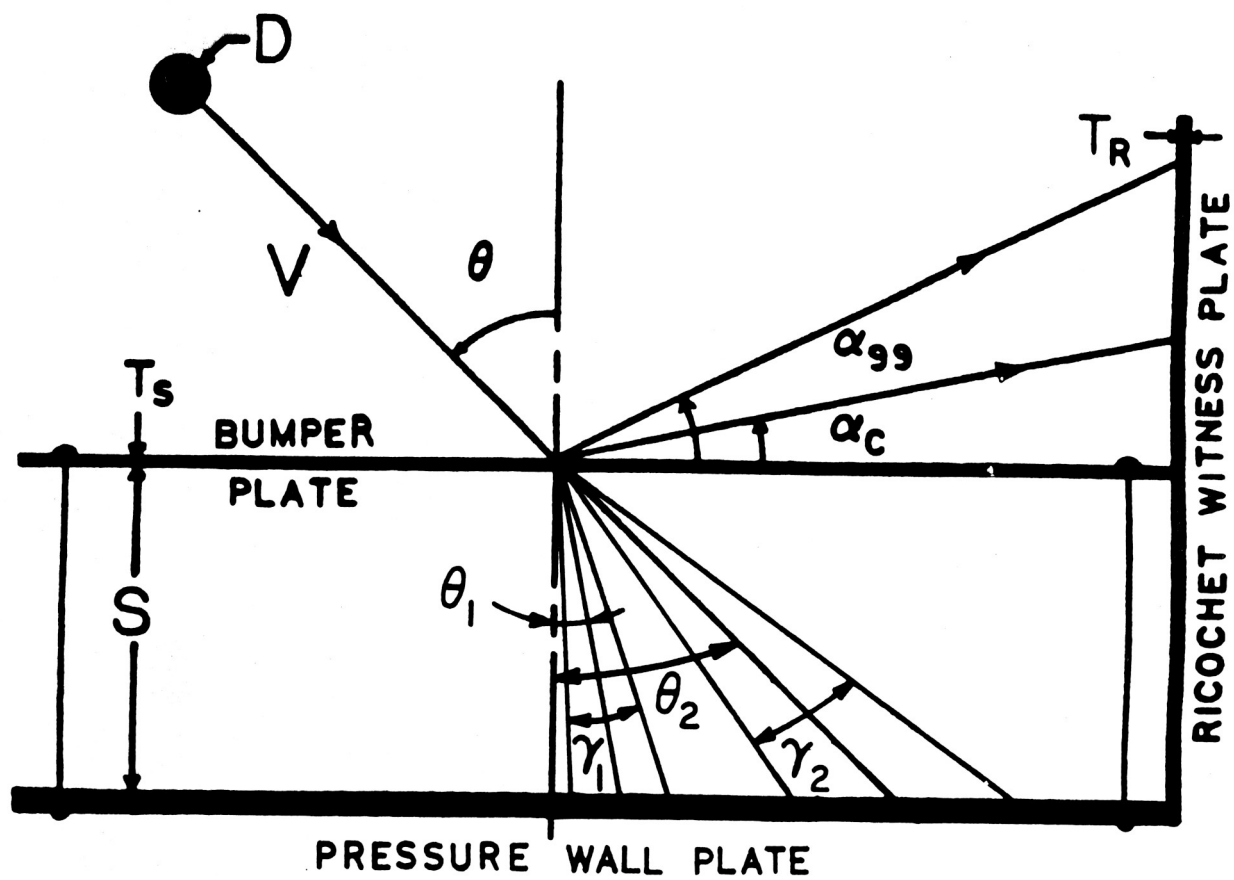


Figure 1 -- Test Configuration and Parameter Definitions

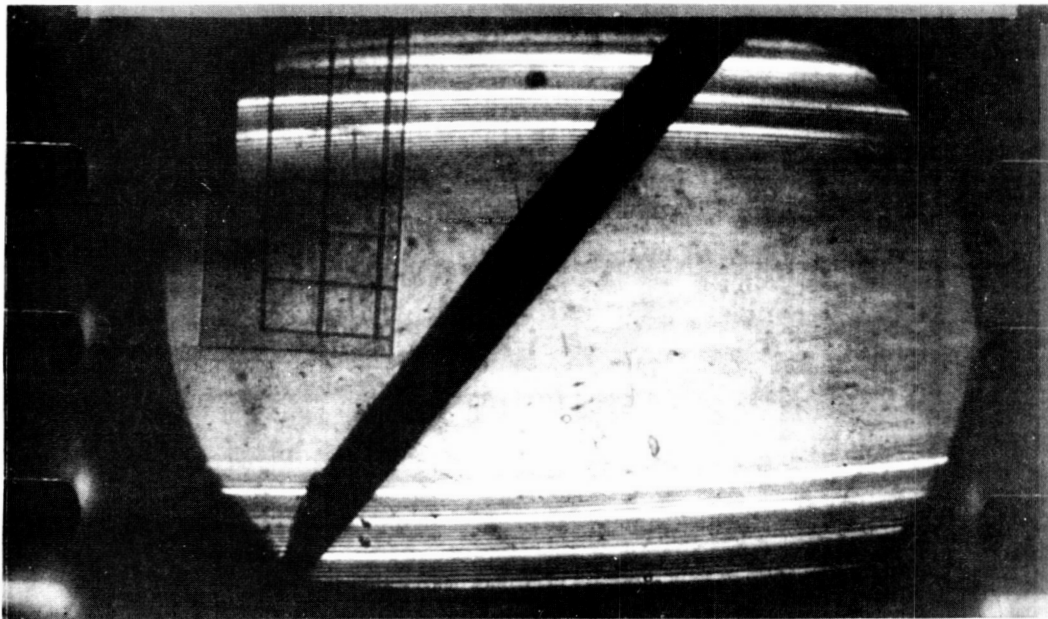


Figure 2 -- Oblique Hypervelocity Impact: Pre-Event Stage
 $V \sim 6 \text{ km/sec}$, $\Theta = 45 \text{ degrees}$, $t_s/d \sim 3$

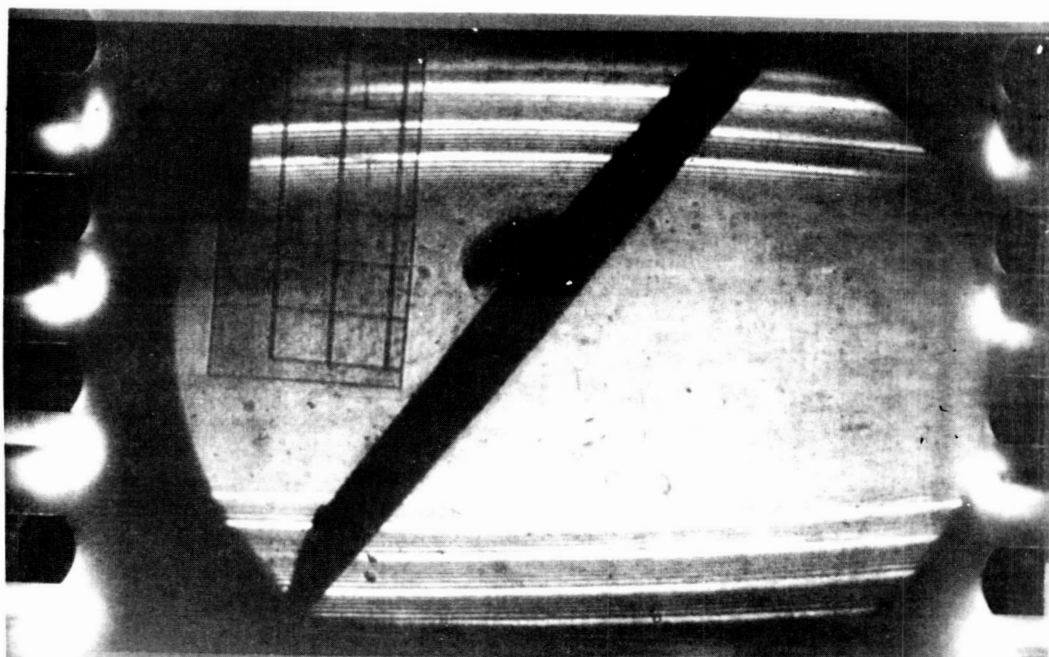


Figure 3 -- Oblique Hypervelocity Impact: Formation
 of Ricochet Debris Cloud
 $V \sim 6 \text{ km/sec}$, $\Theta = 45 \text{ degrees}$, $t_s/d \sim 3$

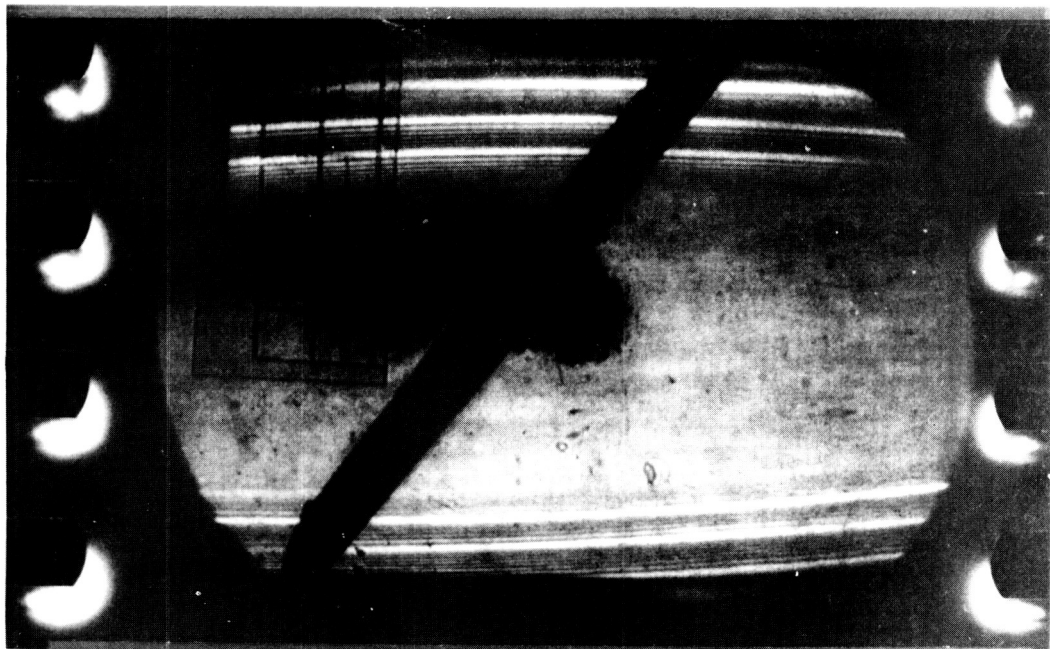


Figure 4 -- Oblique Hypervelocity Impact: Formation
of Penetration Debris Cloud
 $V \sim 6 \text{ km/sec}$, $\Theta = 45 \text{ degrees}$, $t_s/d \sim 3$

ORIGINAL PAGE IS
OF POOR QUALITY

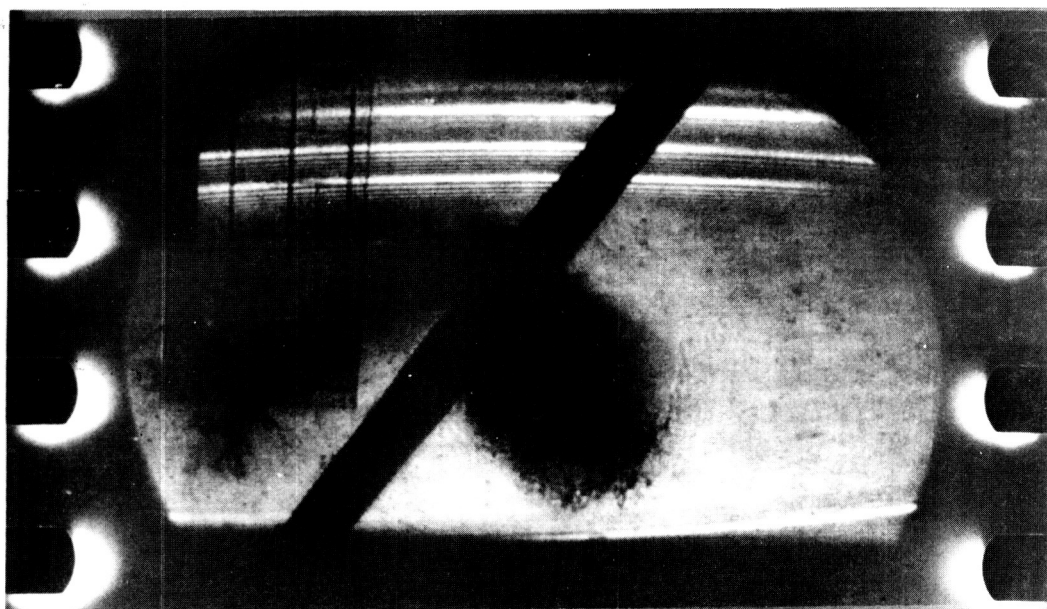


Figure 5 -- Oblique Hypervelocity Impact: Spread of
Ricochet and Penetration Debris Clouds
 $V \sim 6 \text{ km/sec}$, $\Theta = 45 \text{ degrees}$, $t_s/d \sim 3$

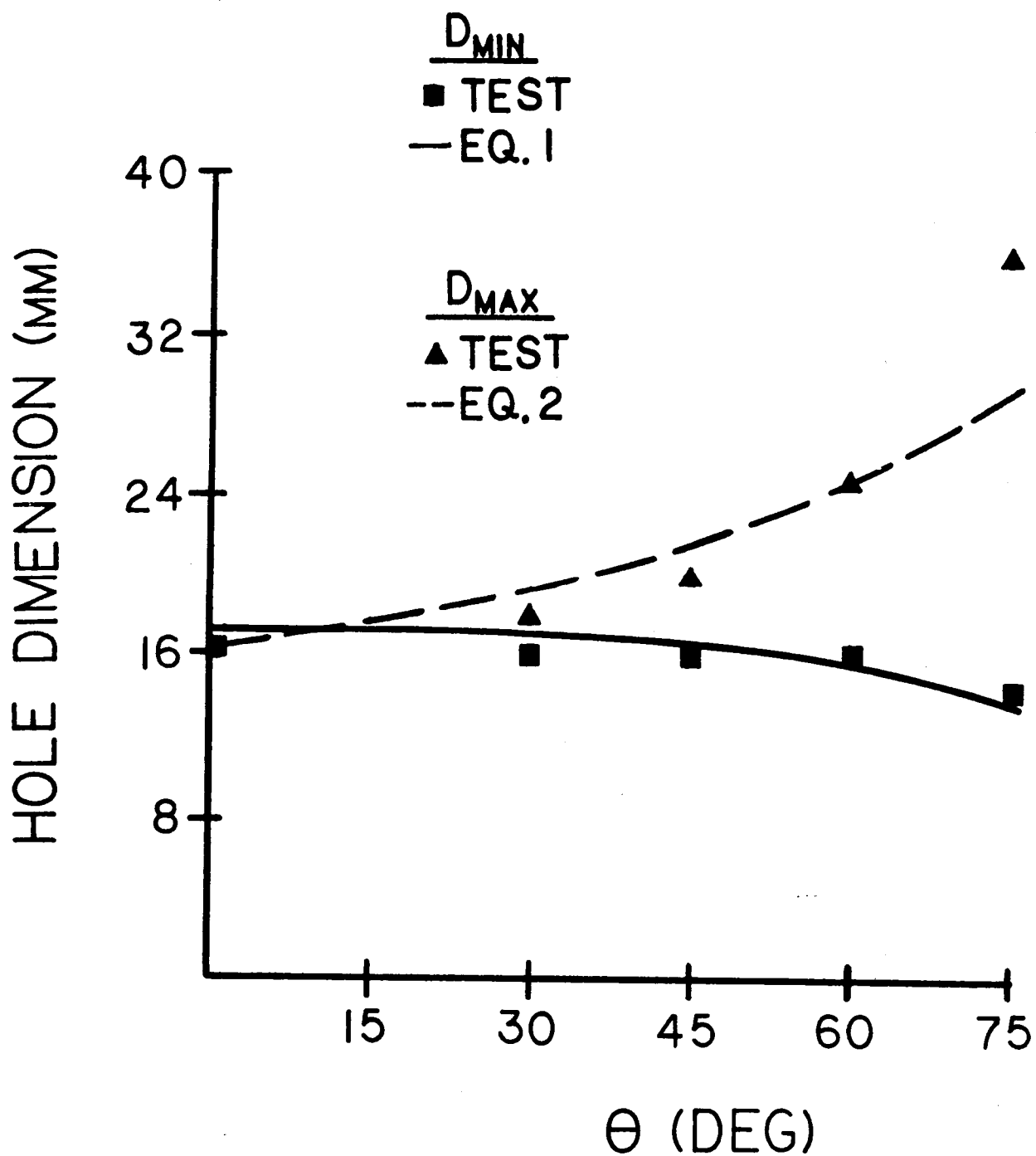


Figure 6 -- Bumper Plate Hole Dimensions: Test Data Compared With Regression Equation Predictions
($V \sim 7$ km/sec, $d = 7.95$ mm)

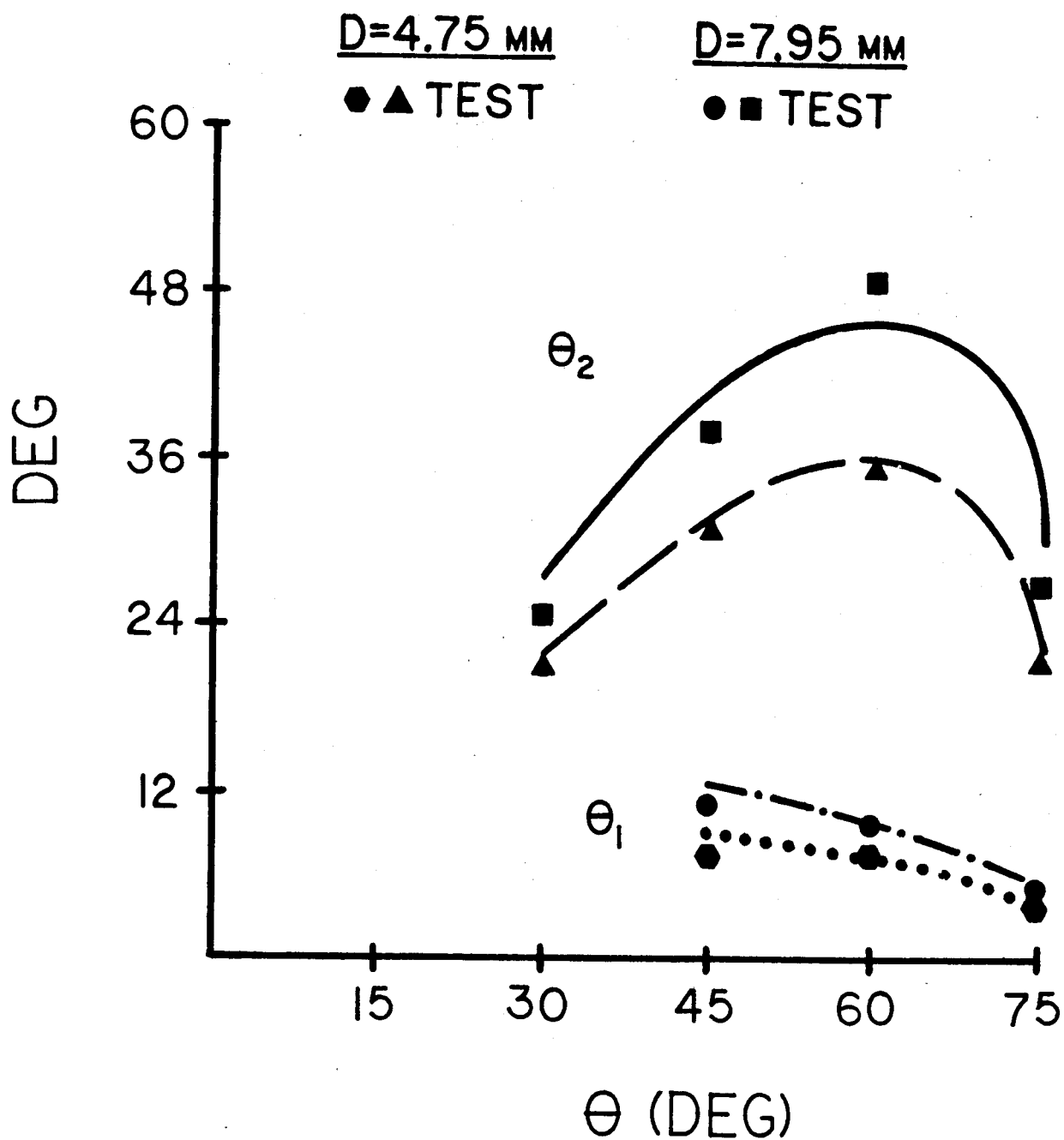


Figure 7 -- Penetration Fragments Trajectories: Test Data Compared With Regression Equation Predictions
($V \sim 7$ km/sec, $d = 7.95$ mm)

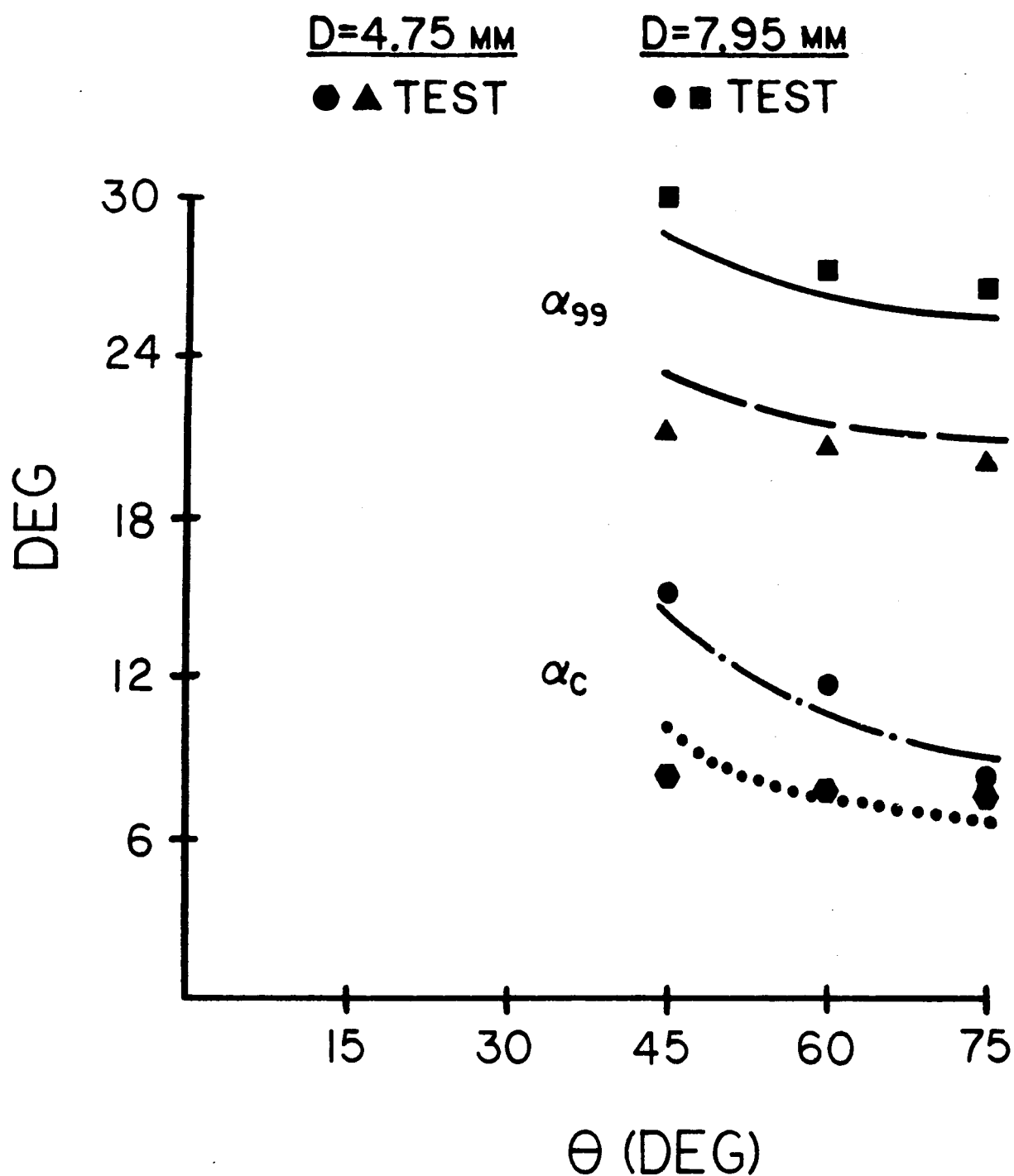


Figure 8 -- Ricochet Fragments Trajectories: Test Data Compared With Regression Equation Predictions
($V \sim 7 \text{ km/sec}$, $d = 7.95 \text{ mm}$)

ORIGINAL PAGE IS
OF POOR QUALITY

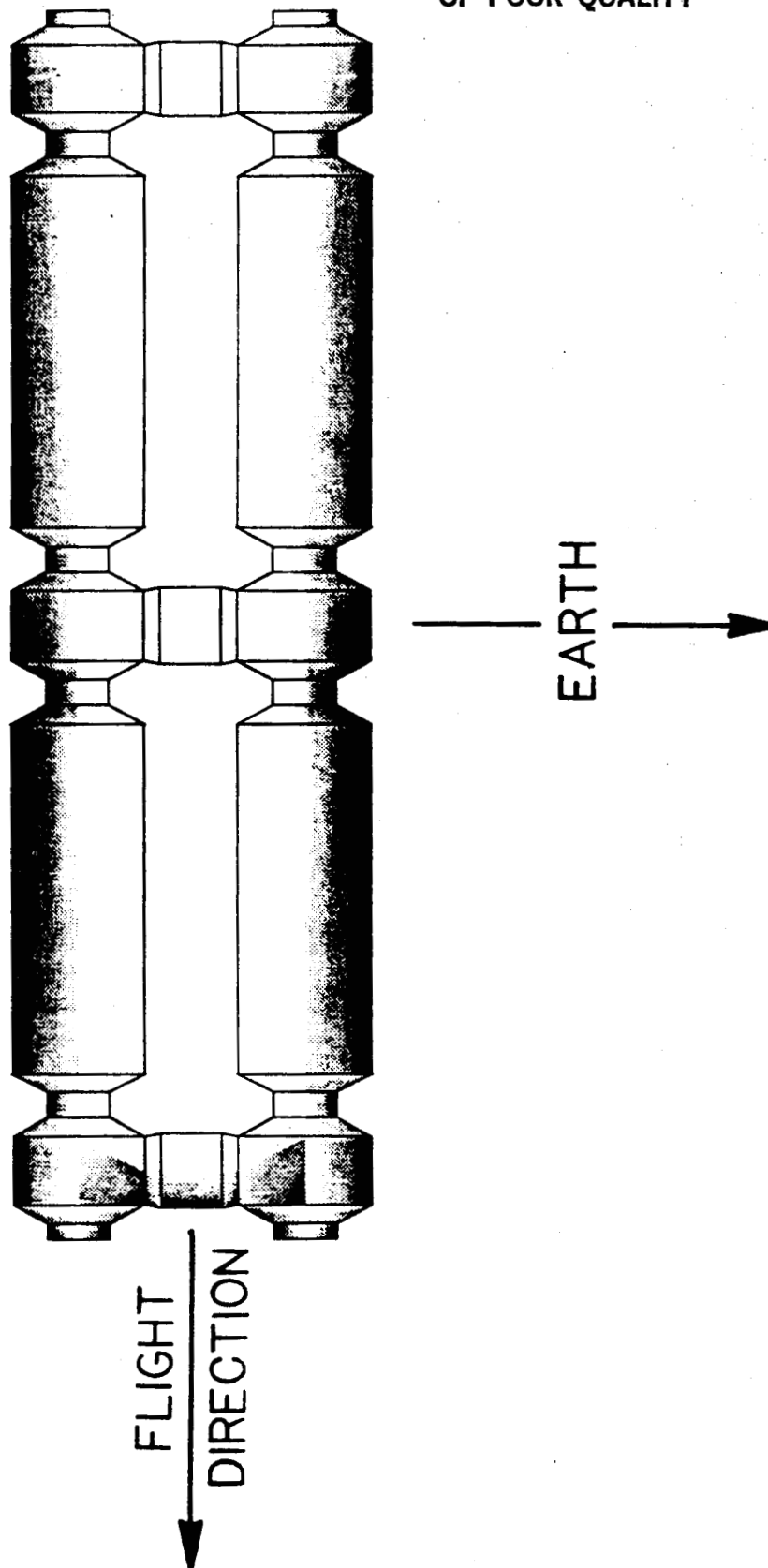


Figure 9 -- Space Station Module Orientation

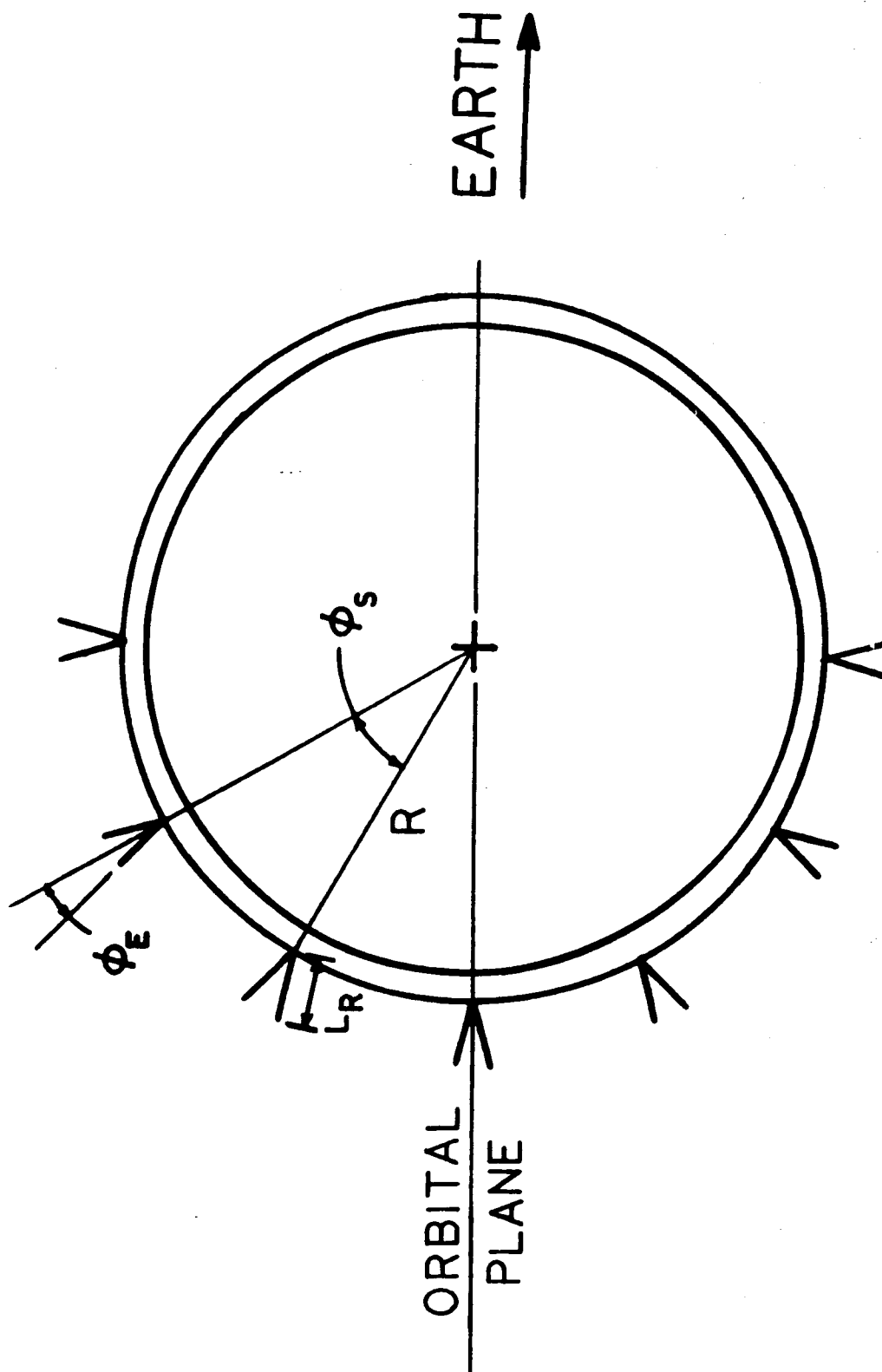


Figure 10 -- Space Station Module Cross-Section
With Containment Shield Parameters

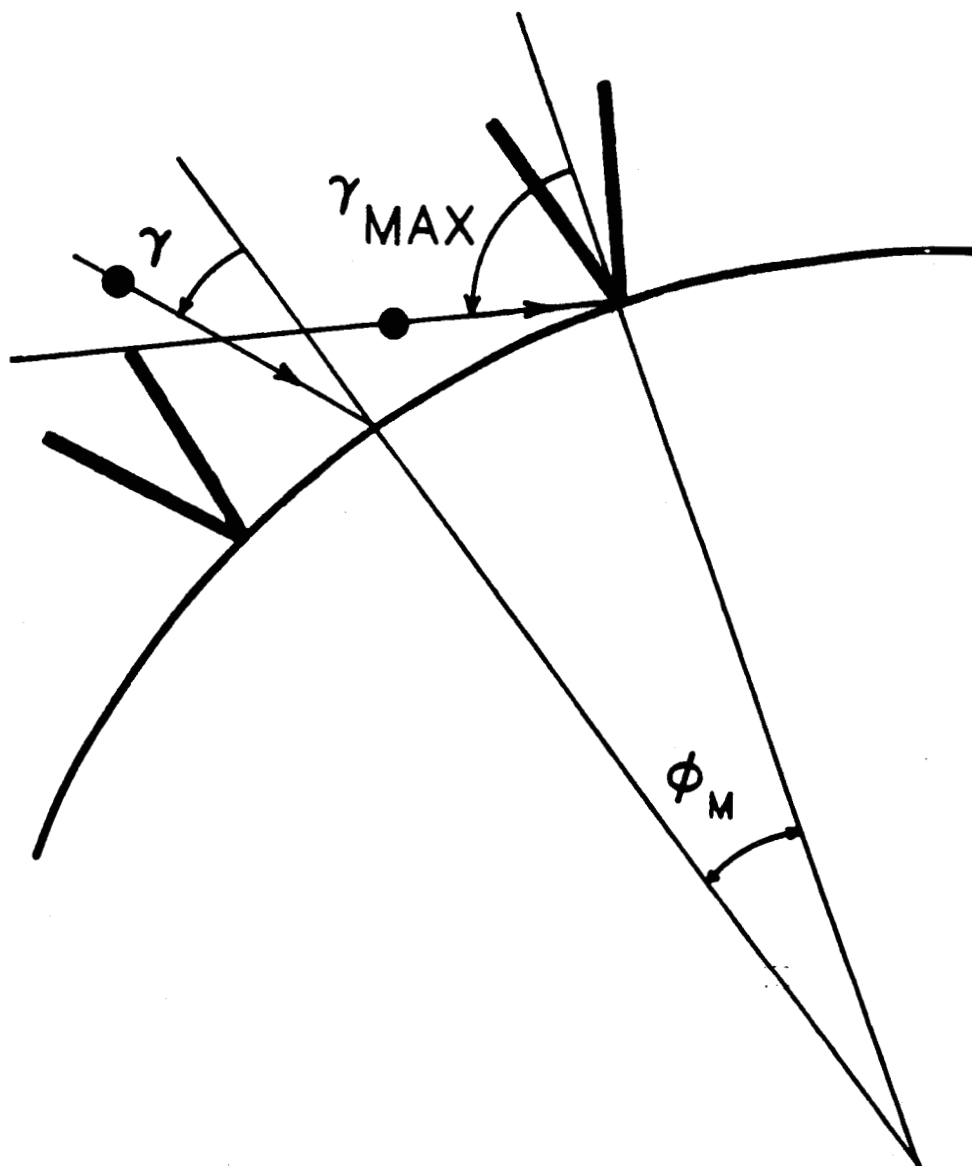


Figure 11 -- Meteoroid/Space Debris Trajectories

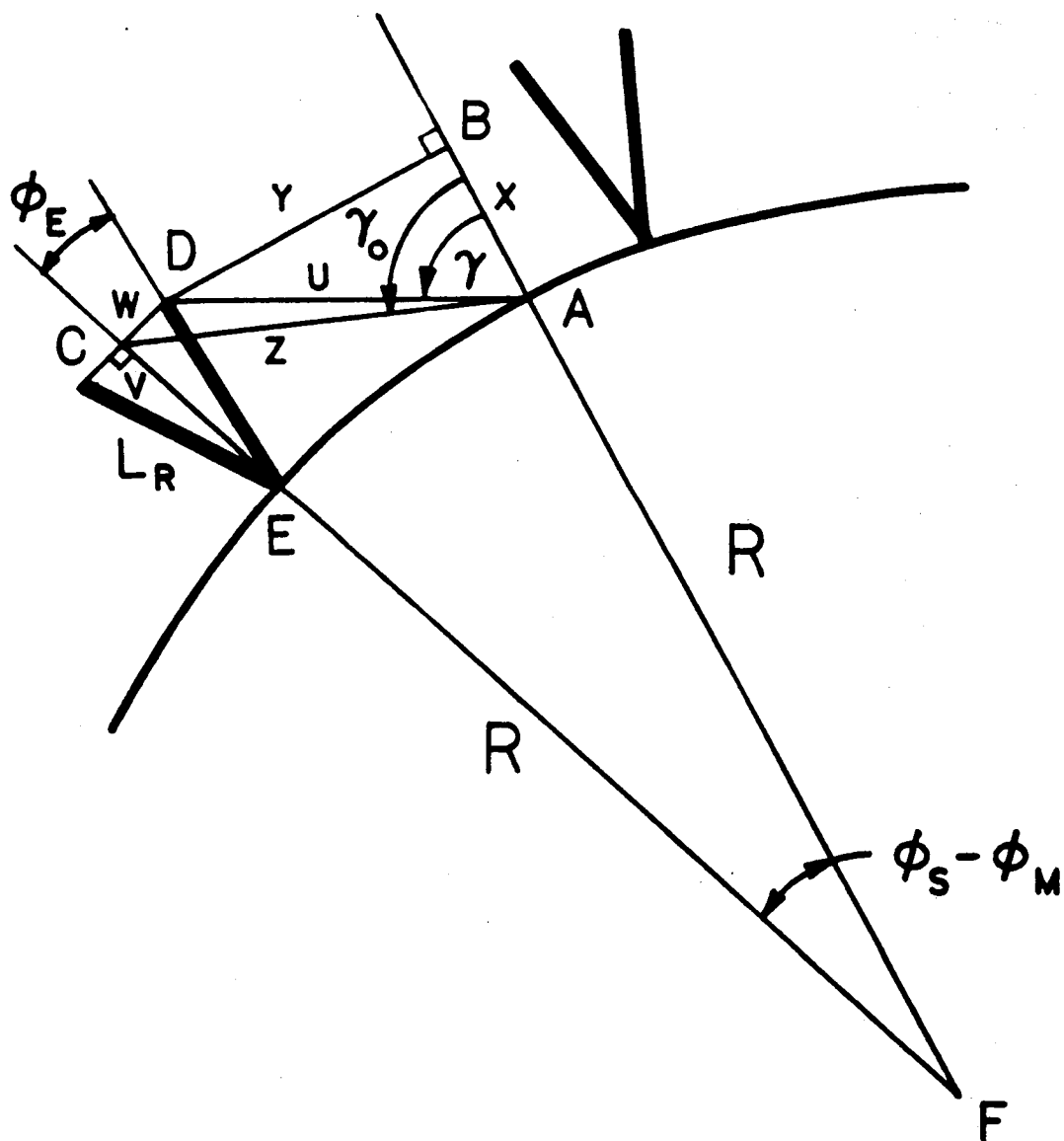


Figure 12 -- Containment Shield Geometry for Design Condition #1

N89-21751

1988

NASA/ASEE SUMMER FACULTY RESEARCH FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

EVALUATE THE APPLICATION OF MODAL TEST AND ANALYSIS
PROCESSES TO STRUCTURAL FAULT DETECTION IN
MSFC - STS PROJECT ELEMENTS

Prepared By: William T. Springer
Academic Rank: Assistant Professor
University and Department: University of Arkansas
Mechanical Engineering

NASA/MSFC:

Laboratory: Test
Division: Structural Test
Branch: Dynamic Test

NASA Colleague: A. D. Coleman
Date: July 29, 1988
Contract No.: NGT 01-002-099
The University of Alabama

ABSTRACT

The Space Transportation System (STS) is a very complex and expensive flight system which is intended to carry very unique payloads into low earth orbit and return. A catastrophic failure of the STS (such as experienced in the 51-L incident) results in the loss of both human life as well as very expensive and unique hardware. One impact of this incident has been to reaffirm the need to do everything possible to insure the integrity and reliability of the STS is sufficient to produce a safe flight.

One means of achieving this goal is to expand the number of inspection technologies available for use on the STS. The purpose of the activity reported here was to begin to evaluate the possible use of assessing the structural integrity of STS components for which Marshall Space Flight Center (MSFC) has responsibility. This entailed reviewing the available literature and determining a low-level experimental program which could be carried out by MSFC and would help establish the feasibility of using this technology for structural fault detection.

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to the NASA/ASEE Summer Faculty Fellowship Program for the opportunity to participate in this program. In addition, he would like to specifically thank Michael Freeman, Ernestine Cothran, and Billie Swinford for the support and guidance they provided this summer.

To NASA Colleagues A. D. Coleman and C. A. Kirby (Chief) of the Dynamics TEST Branch, and G. B. Waggoner, Chief of the Structural Test Division, the author wishes to extend his gratitude for their interest and encouragement.

Finally, to G. D. Johnston (former Chief, Dynamics Test Branch) of Vibration and Acoustics, Inc., thanks are due for his support and assistance, both past and present.

INTRODUCTION

During recent years it has become increasingly necessary to assess the integrity of existing structural systems so as to determine their suitability for continued use. Examples of these structural systems include commercial and military aircraft, offshore oil platforms, commercial buildings in earthquake prone regions, and bridges. All of these examples are structures which represent a significant initial capital investment and/or the potential for harm to either human occupants or the environment or both if they fail.

Recently, the Space Transportation System (STS) has been added to this list. The STS has many components which are reusable, and the Orbiter Vehicle is the most recognizable of these. In addition to the usual problems encountered in inspecting and evaluating the condition of a typical aerospace structure, the STS has several unique requirements which must be accommodated. These include exposure to an extremely hostile and wide ranging environment, the need to significantly reconfigure the vehicle for each flight, and a very tight schedule of access to the vehicle between flights.

The technology currently used to assess the integrity of the orbiter and other pieces of flight structure requires direct access to the areas to be inspected. Additionally, these technologies (such as ultrasonics, acoustic emissions, dye penetrants, and others) require the surface of the component to be inspected to be clean, and in many instances, the component must be physically removed from the structure. The result is an extremely localized inspection program which requires the expenditure of large amounts of time and labor, and can result in a degradation of the structural integrity just due to the disassembly-reassembly cycle involved. The DC-10 crash on take-off in Chicago a few years ago is a dramatic example of this possibility.

Less tragic but equally costly results can be produced by simple mistakes such as a bearing being improperly installed after an inspection of a large piece of equipment in the petrochemical or power generation industry. Therefore, any inspection scheme which either eliminates the need for or reduces the frequency of this type of detailed inspection process without sacrificing the quality of the inspection

can provide significant savings to the industry using it.

This approach has been used successfully in many process related industries and is referred to as "preventive maintenance". By monitoring such things as machine vibrations and the chemical composition of lubricants on a regular basis, it is possible to keep a running log of the condition of important pieces of equipment without removing them from service. In this manner, down-time and equipment outages can be controlled and scheduled to coincide with other related activities.

The purpose of the work performed under this program was to begin the evaluation of modal analysis methods for use as a structural fault detection tool. The major advantages this method has over those currently in use are that it is a global inspection technique, and does not require disassembly of the structure to be inspected. It may also be possible to extend the application of this technology to a point where it can function in a preventive maintenance role. However, the current interest in this technology centers around its global measurement capability, and the fact that it can be utilized without requiring any disassembly of the structure.

PROJECT DESCRIPTION

As mentioned earlier, modal analysis and testing techniques have two major benefits to offer as a Non-Destructive Evaluation (NDE) method. These are 1) the global nature of the evaluation and 2) the fact that the evaluation can be performed without disassembling the structure. However, there are some significant questions which must be investigated before it will be possible to determine the extent of the usefulness of this method as an NDE tool.

These concern themselves with such things as the type of fixturing, if any, required, the amount and type of data to be taken, the accuracy and repeatability required for the data, the type of data processing to be used, the manner in which the data should be presented, the rationale for interpreting the data, and so forth. While this list is far from complete, it serves to illustrate the complexity of this issue.

Therefore, it was decided that the best approach to take in the present effort was to review the literature available on this topic, and then begin a low-level test program that could be carried out by the Dynamic Test Branch at MFSC. This test program would serve to familiarize their personnel with the use of modal test methods for NDE purposes, illustrate the impact of accuracy and repeatability on the test data, allow some preliminary evaluation of various data processing and presentation methods, and to generally illustrate how typical aerospace structural systems will behave when damage is present. The results of these preliminary investigations will provide the basic information required to determine how best to proceed with a more advanced and ambitious evaluation program.

Literature Survey

The fact that the frequency content of a vibrating system changes as the physical characteristics of the system change has been used to advantage in several areas. These range from the above mentioned application to preventive maintenance and troubleshooting as described by Downham and Woods [5], Glew and Watson [6], Cempel [82], and Mutch and Russell [86] to the determination of the elastic properties of structural adhesives as explained by Adams and Cappendale [81].

One does not have to expend much effort to discover that this concept has been applied in many different forms to many different problems. This is demonstrated by the set of references listed at the end of this report. While this list is by no means complete, it does serve to illustrate the point quite well. In particular, these articles can be divided into several general categories, and they will be described in this manner.

Two areas which have received considerable attention in recent years are civil engineering structures and offshore oil platforms. Yao and his colleagues [11,15,16,78-80] have published extensively on the subject of inspecting civil engineering structures. The goal being to define a method of assessing the integrity of these structures after their exposure to some situation such as an earthquake which could severely damage a structure without causing it to collapse. The outcome of this work has been a probabilistic method to quantify the damage as either minor or moderate or severe coupled with an attempt to computerize the evaluation process. Inputs range from subjective evaluations made by inspectors to actual empirical data. There is no indication of how successfully the system works.

Krauthammer [9] has suggested a numerical evaluation plan similar in intent to that of Yao. Matzen and Hardee [10] have demonstrated a numerical method based on the Hessian matrix for determining if an element in an FEM model has a value of stiffness equal to zero. The ability of the method to determine a value other than zero was not studied. Sparks, Jeary and de Souza [17] are but one set of investigators that have examined the use of ambient excitation to determine the dynamic properties of a building. Kopff [21] has attempted to set down a rationale for inspecting large natural convection cooling towers also using ambient excitation. All of these investigations are lacking both experimental verification and a method of correlating the presence of damage to changes in dynamic properties.

Another type of structure which has received a fair amount of attention is the offshore oil platform. This class of structure is both unique and extremely expensive. Also, a sudden failure of such a structure can have a tremendous impact on the environment as well as the potential loss of human life. The two main obstacles faced by investigators in this area are both related to economics. The first is the general need to find an answer quickly when structural damage is suspected coupled with the lack of interest on the part of the petroleum industry in supporting fundamental

research on this topic. The second is the fact that the platform cannot be casually taken out of service at any time during its life. The cost of putting these structures into service and then keeping them operating is enormous. Therefore, any NDE which must be performed on these structures must take place while the platform is operating. This can complicate the measurement and analysis process to the point where it is virtually fruitless.

Loland, Mackenzie, Begg and Bendat [8,14,19] have highlighted some of these difficulties. One of the most significant of these is determining an adequate model of the platform which can be used with modal analysis procedures to determine structural integrity. This is not a simple problem. Kummer, Yang and Dagalakis [13] have applied the "log decrement" technique to simple structures to detect the presence of fatigue cracks with an eye to applying the results to offshore platforms. This application to real structures has yet to be accomplished. Vanduerzen, Leuridan and Doucet [28] have also applied modal analysis techniques to an offshore platform. Again, the model of the platform used was extremely simple and made it difficult to determine the precise condition of the structure.

Large rotating equipment has also received attention. The concern in these papers is the ability to avoid an unexpected catastrophic failure of these extremely large and expensive pieces of equipment. In addition, if the need to disassemble this equipment unnecessarily can be avoided, the financial benefits to the owner are significant. This concern results from the fact that the materials used in these large pieces of equipment are not, typically, very damage tolerant. Therefore, a relatively small crack can be devastating. References 59 to 63 and 67 describe several different variations of this problem. In all cases, no clear choice for an obvious monitoring scheme is identified.

There has also been much written in way of what can be termed general application of vibration analysis to damage detection or the detection of structural changes. The amount of published literature is extensive and References 20, 22, 23, 25-27, 29, 30, 43-58, 64-72, 84, 85 and 88 constitute a small but representative sample of the work in this area. Virtually all of this work concentrates on examining very simple structures or structural elements and, again, it is unclear whether this information can be applied to real structural systems.

There have also been a number of theses and dissertations written on this general subject, and References 31 to 41 are

reflective of this work. Of this body of work, only the contributions of Haisty [40] and Duerr [41] have addressed the need of providing the analyst as well as the experimentalist with the capability of evaluating the effect damage will have on the dynamic properties of a structure. In the long term, this will be an important feature of this technology. Again, none of this work has been applied to large structural systems.

One method which seems to be finding some favor is to force an FEM model to exhibit the measured structural vibration characteristics. Typically, this approach uses some type of optimization routine to determine the final configuration. Most of the work which falls into this category is either proprietary or else just developing. More of this work should be coming to publication in the near future.

Present Investigation

The need to monitor the structural integrity of aerospace structures is well established. The number of these systems in commercial service alone is staggering. Recent estimates put the number of departures and arrivals in excess of 26,000 per day. Each of these flights carries a large number of passengers and operates, at least a portion of the time, in areas of high population density. These facts all combine to produce a situation that would result in the loss of a large number of human lives if a catastrophic structural failure should occur.

In addition to this pressure, the military also has a tremendous inventory of aircraft of all types. While, in general, these aircraft carry fewer passengers and crew than those in civilian service, the sheer number of aircraft in the military inventory establishes an economic imperative, if nothing else, to keep them in the best possible health. However, it is not at all clear how the needed inspections should be accomplished. This situation is well documented by Sproat and Rowe [87]. They point out how the existence of all the competing inspection technologies complicate the issue even more. This reference details a four year test program carried out by Lockheed-Georgia for the Air Force which was intended to evaluate all of these competing methods. But, as stated earlier, everyone agrees that inspections are needed.

One approach to this problem has been to measure the strain experienced by the structural component or assembly of interest and use this data to make a determination of its integrity. An early attempt to implement this strategy is

documented by Bochniak and Garcia [42]. This report describes a twelve channel strain measurement and recording system which was developed for the Navy, and intended for use as an aircraft structural integrity monitoring device. There is no information concerning the success of the monitoring program.

A more recent attempt to accomplish the same goal is presented by Bruinsma et al. [3], and details the use of fiber-optic strain measurements to provide information concerning the structural integrity of composite materials. Again, this reference describes only the data acquisition strategy and no information is presented concerning an actual application of this device in a real system.

The investigation presented here resulted, in part, from the interest generated by a Level II Change Request and PRCBD #S40155 titled "Orbiter Structural Inspection Via Modal Processes." This request was initiated by Mr. Walter M. West Jr., JSC-ES4, as a possible solution to the problem of inspecting components of the shuttle which cannot practically be inspected using other technologies. The particular items of concern are control surfaces. These components are typically built-up structures and can sustain internal damage which cannot be detected by conventional NDE methods without removal of the component's aluminum skin.

The prime motivation for this concern resulted from an acoustic test of the body flap of the Space Shuttle. Cheng, Dunham and Joanides [83] have documented the basic test procedure and expected test results. After the completion of this evaluation program on the body flap, a modal survey revealed that some of its dynamic characteristics had changed. West determined that damage had accumulated in the body flap by reviewing the vibrational mode shapes.

Later, West [74-76] developed a technique for locating damage in structures using a modal vector correlation coefficient. The structure is partitioned into several components and the correlation coefficient between baseline and post-service mode shapes are calculated. For components in which no change has occurred, the correlation coefficient is near unity, while in damaged segments the value of the correlation coefficient is less than one. The procedure was applied to Space Shuttle test specimens with success.

Haisty, West and Mitchell [77] applied West's technique to the Space Shuttle Orbiter body flap test specimen with good results. Several damage sites were correctly identified using this technique which were not detected by conventional

NDE techniques. While this technique can locate damage in structures reasonably well, it is unable to determine the extent of the damage.

As an extension of these activities, work has recently completed at JSC which applied this technique to a small Cessna airframe. The goal was to inspect the control surfaces while they are installed in the airplane. The initial damage placed in the structure was used to evaluate the concept of using local modes (i.e. modes primarily attributed to the subsystem only) to detect damage in the control surfaces. This work has produced results which were very favorable and has led to the inclusion of this approach in the inspection program for the Orbiter.

Further support for the use of modal analysis procedures as an NDE tool has been provided by the Structural Dynamics Research Corporation [73]. SDRC has reported cases in which damage in spacecraft and airplanes has been diagnosed as the result of routine modal tests. They conclude that modal analysis can be readily adapted to the identification of structural failures, that frequency response functions can quickly give an indication of whether or not damage has occurred in the structure, and the mode shapes can provide information about where the structural problem has occurred. This work also stops short of quantifying the damage present.

The preliminary results obtained in the experimentation begun this summer at MSFC-ET53 demonstrates that the success achieved at JSC may not translate directly to all structures. The boxbeam structure tested in this program was damaged in both primary and secondary structural members and the MAC values calculated in the same manner as was done at JSC. The results indicate that for this structure at least the damage does not become apparent in the MAC values until rather large cuts have been made. Also, the modes which intuitively seem like they would be the most effected are not the modes which exhibit the greatest variation in MAC values. This gives further credence to the need for more work in this area so that a logical strategy for applying this technology can be formulated.

In addition to the work completed at JSC and the interest in evaluating this technology for use by MSFC-ET53, there are additional groups within NASA with similar problems. The first of these is MSFC-PD22. This group is working on development programs utilizing new materials which will have very unique properties. Because of this, it will be extremely important to have reliable, efficient, and

economic NDE methods available for use on these systems. The propulsion group at MSFC also has interest in using measured quantities to determine the condition of the SSME and other related hardware. Many of the problems associated with these activities are similar, at least in form, to those reviewed in this report. Last, there is an apparent interest at the Langley Research Center in evaluating the application of modal analysis methods to the problem of assessing acoustic fatigue damage.

Existing Problems

The following problems currently exist and must be solved in order to make modal analysis an effective and useful NDE method:

1. Measurement noise.
2. Available excitation techniques for certain situations may limit the usefulness of the data.
3. Beat methods for data presentation and processing are unclear.
4. Measurement accuracy and repeatability requirements are unknown.
5. Interpretation of data to yield damage level and/or location values not yet clearly possible in real structures.
6. Quantity of data required to make an accurate damage assessment is unknown.
7. Low cost transducer development is needed.
8. Very little understanding of damage/structure interaction.
9. Much information on this general topic is available in the literature but its usefulness is not clear.
10. Expense of hardware and software required for data acquisition and analysis is still high.
11. Desire to apply this technology to NDE of new materials before simple structures are understood.
12. Uniqueness of the predicted damage location and/or level values may be questionable.

13. Fixturing needed to insure reproduceable test conditions is unclear and maybe extensive.

The problem of measurement noise is being addressed by Wicks and Mitchell [1,2,24], and their work seems to be bringing this problem under control. Concern over the available excitation methods is of less importance in the aerospace industry than in the other application areas mentioned earlier because of the control one has over the test article during the inspection process.

In the future, on the other hand, if this method is to be expanded to include in-service responses in the evaluation process as well, the information contained in these measurements will have to be examined very closely to insure that it will produce useful estimates of the integrity of the structure. It is possible that the measurement sites selected for use in the modal test which would be done when the system is out of service may not provide the best possible information from the in-service measurements.

Another issue is the number of in-service measurement sites needed. This number will have a significant impact on not only the quality of the data but on the cost of the monitoring system as well. Therefore, it will be very important to select the measurement sites with great care and deliberation.

The need to develop low cost transducers stems from the desire to eventually be able to install the required instrumentation in the vehicle as a part of the construction process. In this way considerable effort and expense can be avoided during the service life of the vehicle by eliminating the need to install and remove large numbers of transducers each time a structural integrity assessment was to be made. At the same time, adequate instrumentation would still be available to permit a complete inspection to be performed.

Obviously, how this technology is applied by the general public and the cost of the data acquisition and analysis equipment are items which cannot be controlled. However, these are two issues which will play a significant role in the development and utility of this technology for NDE purposes.

The remainder of the items listed above will have to be addressed through comprehensive research efforts.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The ability of modal analysis methods to provide information which can be used for the NDE of structural systems is clear. However, many issues concerning the efficient and accurate utilization of this data for NDE purposes are unresolved. These include, but are not limited to, the amount of data needed, how accurate and repeatable the data must be, how this data should be processed and presented, and how the damage interacts with the dynamic characteristics of the structure. A comprehensive research effort will be required to successfully address these concerns.

Recommendations

The preliminary results obtained this summer indicate both the promise and pitfalls to be encountered in using modal analysis methods for NDE of structural systems. Based on these results, the recommendations listed below will provide needed insight into the problem areas delineated earlier.

1. Begin a research effort which will determine which of the various dynamic characteristics should be monitored, provide guidance on how to monitor them, and give an indication of the measurement accuracy needed when they are monitored.
2. Begin a research effort to determine if it will be possible to characterize as well as locate the structural faults of interest in a particular structure.
3. Investigate how to incorporate this ability into the structural design/analysis cycle so that future designs can be improved.
4. Begin a basic research effort to characterize the dynamic behavior of fundamental composite structural elements in the same fashion as has been done for metallic structures.

BIBLIOGRAPHY

1. Wicks, A. L. and Mitchell, L. D., "A New Means of Estimation of Frequency Response Functions in the Presence of Uncorrelated Signals, Part I," Proceedings of the Spring Meeting of the Society for Experimental Mechanics, 1987, pp. 629-634.
2. Wicks, A. L., "Frequency Response Function Estimation and Quality Statistic," Proceedings of the Spring Meeting of the Society for Experimental Mechanics, 1987, pp. 517-523.
3. Bruinsma, A. J. A., et al., "Fibre-Optic Strain Measurement for Structural Integrity Monitoring," Proceedings of the Second International Conference on Optical Sensors, 1984, pp. 399-402.
4. Rakshin, A. F. and Sidorov, O. T., "Determination of the Degree of Defectiveness in Structural Components," Soviet Journal of Nondestructive Testing, Vol. 18, No. 3, November 1982, pp. 158-161 (Translation).
5. Downham, E. and Woods, R., "The Rationale of Monitoring Vibration on Rotating Machinery in Continuously Operating Process Plant," ASME Paper No. 71-Vibr-96, 1971.
6. Glew, C. A. W. and Watson, D., "Vibration Analysis as a Maintenance Tool," Transactions of the Institute of Marine Engineers, Canadian Division, Supp. No. 32, June 1968, pp. 18-29.
7. Petroski, H. J. and Glazik, J. L., Jr., "The Response of Cracked Cylindrical Shells," Journal of Applied Mechanics, Trans. ASME, Vol. 47, June 1980, pp. 444-446.
8. Loland, O. and Mackenzie, A. C., "On the Natural Frequencies of Damaged Offshore Oil Platforms," Mech. Res. Comm., Vol. 1, 1974, pp. 353-354.
9. Krauthammer, T., "A 'Numerical Gauge' for Structural Assessment," Shock and Vibration Bulletin, No. 56, Part 1, 1986, pp. 179-194.
10. Matzen, Vernon Charles and Hardee, Joseph Earlton, Jr., "Mathematical Modelling of Indeterminate Trusses," Proceedings of the Second Specialty Conference on Dynamic

Response of Structures, 1981, pp. 687-697.

11. Yao, J. T. P., "Damage Detection Using System Identification," Proceedings of the Spring Meeting of the Society for Experimental Mechanics, 1987, pp. 383-384.

12. Liu, Dahsin, Sun, C. T. and Malvern, L. E., "Structural Degradation of Impacted Graphite/Epoxy Laminates," Shock and Vibration Bulletin, No. 56, Part 2, 1986, pp. 51-60.

13. Kummer, E., Yang, J. C. S. and Dagalakakis, N., "Detection of Fatigue Cracks in Structural Members," Proceedings of the Second Speciality Conference on Dynamic Response of Structures, 1981, pp. 445-461.

14. Loland, O., Mackenzie, A. C. and Begg, R. D., "Integrity Monitoring of Fixed, Steel Offshore Oil Platforms," BSSM/RINA Joint Conference of Measurement in the Offshore Industry, 1975.

15. Liu, Shih-Chi and Yao, James T. P., "Structural Identification Concept," Journal Of The Structural Division, ASCE, Vol. 104, ST12, December 1978, pp. 1845-1858.

16. Yao, James T. P., "Damage Assessment of Existing Structures," Journal Of The Engineering Mechanics Division, ASCE, Vol. 106, No. EM4, August 1980, pp. 785-799.

17. Sparks, P. R., Jeary, A. P. and de Souza, V. C. M., "A Study of the Use of Ambient Vibration Measurements to Detect Changes in the Structural Characteristics of a Building," Proceedings of the Second Speciality Conference on Dynamic Response of Structures, 1981, pp. 189-200.

18. Cole, Henry A., Jr., "On-Line Failure Detection and Damping Measurements of Aerospace Structures by Random Decrement Signatures," NASA CR-2205, 1973, 75 pages.

19. Begg, R. D. and Bendat, J. S., "Instrumentation and Analysis Techniques for the Monitoring of Offshore Structures," Petroleum Times, October 1, 1976, pp. 29 and 35.

20. Wang, J. T. S., Liu, Y. Y. and Gibby, J. A., "Vibrations of Split Beams," Journal of Sound and Vibration, Vol. 84, No. 4, 1982, pp. 491-502.

21. Kopff, Paul, "Experimental Modal Analysis for Structural Damage Assessment: The Case of Natural Convection Cooling Towers," Proceedings of the Fifth International Modal

Analysis Conference, 1987, pp. 122-126.

22. Lai, Hsin-Yi, "Computer Aided Fault Diagnosis of Crank Systems Using Engine Vibration Data," Proceedings of the Fifth International Modal Analysis Conference, 1987, pp. 248-253.

23. Brown, T. A., "The Identification of Spatial Parameter Changes Via Modal Analysis," Proceedings of the Fifth International Modal Analysis Conference, 1987, pp. 267-273.

24. Mitchell, L. D., et al., "An Unbiased Frequency Response Function Estimator," Proceedings of the Fifth International Modal Analysis Conference, 1987, pp. 364-373.

25. Hopton, Gregory W. and Deblauwe, Filip, "Relationships Between Analysis Errors and Complex Mode Shapes," Proceedings of the Fifth International Modal Analysis Conference, 1987, pp. 381-388.

26. Afolabi, D., "An Anti-Resonance Technique for Detecting Structural Damage," Proceedings of the Fifth International Modal Analysis Conference, 1987, pp. 491-495.

27. Wang, Weiji and Zhang, Azhou, "Sensitivity Analysis in Fault Vibration Diagnosis of Structures," Proceedings of the Fifth International Modal Analysis Conference, 1987, pp. 496-501.

28. Vandeuren, U., Leuridan, J. and Doucet, Y., "Structure Monitoring Using a Diagnosis Technique Based on Combined Use of FEA and Test," Proceedings of the Fifth International Modal Analysis Conference, 1987, pp. 1338-1346.

29. Springer, W. T., Lawrence, K. L. and Lawley, T. J., "The Effect of a Symmetric Discontinuity on Adjacent Material in a Longitudinally Vibrating Uniform Beam," Experimental Mechanics, SEM, To appear.

30. Springer, W. T., Lawrence, K. L. and Lawley, T. J., "Damage Assessment Based on the Structural Frequency Response Function," Experimental Mechanics, SEM, To appear.

31. Bourne, C. A., Material Damping As A Means Of Quantifying Fatigue Damage, Masters Thesis, Air Force Institute of Technology, 1978.

32. Chen, S. H., System Identification And Damage Assessment Of Existing Structures, Ph.D. Dissertation,

Purdue University, 1980.

33. Toussi, S., System Identification Methods For The Evaluation Of Structural Damage, Ph.D. Dissertation, Purdue University, 1982.

34. Stephens, J. E., Structural Damage Assessment Using Response Measurements, Ph.D. Dissertation, Purdue University, 1985.

35. Tsai, Wen-Hu, Detecting And Locating Damages In Complex Structures By The System Identification Technique, Ph.D. Dissertation, University of Maryland, 1985.

36. Springer, W. T., The Vibrational Characteristics of Uniform Bars Containing a Symmetric Discontinuity in Stiffness Applied to Structural Integrity Monitoring, Ph.D. Dissertation, University of Texas at Arlington, 1982.

37. Reznicek, M. E., The Transverse Vibrational Characteristics of Symmetrically Damaged Bars, Masters Thesis, University of Arkansas, 1984.

38. Haisty, B. S., Damage Assessment Using the Longitudinal Vibrational Characteristics of Uniform Bars, Masters Thesis, University of Arkansas, 1984.

39. Moshrefi, N., The Transverse Vibrational Behavior of a Peripherally Cracked Pipe and a Uniformly Corroded Pipe, Masters Thesis, University of Arkansas, 1985.

40. Haisty, B. S., Development and Application of a Cracked-Beam Finite Element and a Cracked-Beam Modal Element for Assessing the Effects of Damage on the Modal Properties of Mechanical Systems, Ph.D. Dissertation, University of Arkansas, 1986.

41. Duerr, R. R., Development and Evaluation of a Dynamic Transfer Matrix for a Damaged Uniform Beam, Masters Thesis, University of Arkansas, 1987.

42. Bochniak, J. and Garcia, R., Structural Integrity Monitoring System, Report No. NADC-77101-30, 1979, 82 pages.

43. Hetenyi, M., "Deflection of Beams of Varying Cross Section," Trans. ASME, Vol. 59, March 1937, pp. A-49-A-52.

44. Kirmser, P. G., "The Effect of Discontinuities on the Natural Frequencies of Beams," Proceedings of the ASTM, Vol. 44, pp. 897-904.

45. Thomson, W. T., "Vibration of Slender Bars With Discontinuities in Stiffness," Journal of Applied Mechanics, Trans. ASME, Vol. 6, June 1949, pp. 203-207.
46. Adams, R. D., et al., "A Vibration Technique for Non-Destructively Assessing the Integrity of Structures," Journal of Mechanical Engineering Science, Vol. 20, No. 2, 1978, pp. 93-100.
47. Cawley, P. and Adams, R. D., "The Location of Defects in Structures From Measurements of Natural Frequencies," Journal of Strain Analysis, Vol. 14, No. 2, 1979, pp. 49-57.
48. Cawley, P. and Adams, R. D., "Defect Location in Structures by a Vibration Technique," ASME Paper No. 79-DET-46, 1979.
49. Ju, F. D., et al., Diagnosis of Fracture Damage in Simple Structures, Bureau of Engineering Research, Report No. CE-62(82) AFOSR-993-1, October 1982.
50. Ju, F. D., Wong, E. T. and Lopez, T. L., "Modal Method in Diagnosis of Fracture Damage in Simple Structures," Productive Applications of Mechanical Vibrations, Applied Mechanics Division, Vol. 52, ASME, 1982, pp. 113-125.
51. Gudmundson, P., "Eigenfrequency Changes of Structures Due to Cracks, Nothches or Other Geometrical Changes," Journal of Mech. Phys. Solids, Vol. 30, NO. 5, 1982, pp. 339-353.
52. Gudmundson, P., "Changes in Modal Parameters Resulting From Small Cracks," Proceedings of the Second International Modal Analysis Conference, 1984, pp. 690-697.
53. Gudmundson, P., "The Dynamic Behavior of Slender Structures With Cross-Sectional Cracks," Journal of Mech. Phys. Solids, 1983.
54. Reznicek, M. E. and Springer, W. T., "Damage Assessment of Transversely Vibrating Uniform Beams Containing a Symmetric Discontinuity," Proceedings of the Spring Meeting of the Society for Experimental Mechanics, 1985, pp. 404-409.
55. Haisty, B. S. and Springer, W. T., "Identification of Multiple Discontinuities in a Uniform Beam Using Longitudinal Vibration Properties," Proceedings of the 40th Mechanical Failures Prevention Group, National Bureau of Standards, 1985.

56. Haisty, B. S. and Springer, W. T., "The Longitudinal Vibration Characteristics of a Uniform Bar Containing Two Symmetric Discontinuities," Proceedings of the Spring Meeting of the Society for Experimental Mechanics, 1985, pp. 389-393.
57. Moshrefi, N., Sreshta, H. A. and Springer, W. T., "The Transverse Vibration Characteristics of an Externally Damaged Pipe," Vibration Analysis to Improve Reliability and Reduce Failure, Applied Mechanics Division, Vol. , ASME, 1985, pp. 23-30.
58. Al-Ansary, M. D. and Azagem, K. M., "A Modal Model for Fatigue Crack Non-Destructive Testing," Proceedings of the Fourth International Modal Analysis Conference, 1986, pp. 202-208.
59. Henry, T. A. and Okah-Avae, B. E., "Vibrations in Cracked Shafts," Proceedings of the First International Conference of Vibrations in Rotating Machinery, I. Mech. E., Vol. 9, No. C162/76, 1976, pp. 15-19.
60. Mayes, J. W. and Davies, W. G. R., "The Vibrational Behaviour of a Rotating Shaft System Containing a Transverse Crack," Proceedings of the First International Conference of Vibrations in Rotating Machinery, I. Mech. E., Vol. 9, No. C168/76, 1976, pp. 53-64.
61. Gasch, R., "Dynamic Behavior of a Simple Rotor With a Cross-Sectional Crack," Proceedings of the First International Conference of Vibrations in Rotating Machinery, I. Mech. E., Vol. 9, No. C178/76, 1976, pp. 123-128.
62. Grabowski, B., "Vibrational Behavior of a Turbine Rotor Containing a Transverse Crack," Journal of Mechanical Design, Trans. ASME, Vol. 102, No. 1, January 1980, pp. 140-146.
63. Rauch, A., "Shaft Cracking Supervision of Heavy Turbine Rotors by FEM Method," Proceedings of the Third International Modal Analysis Conference, 1985, pp. 714-722.
64. Chondros, T. G. and Dimarogonas, A. D., "Identification of Cracks in Circular Plates Welded at the Contour," ASME Paper No. 79-DET-106, 1979.
65. Chondros, T. G. and Dimarogonas, A. D., "Identification of Cracks in Welded Joints of Complex Structures," Journal of Sound and Vibration, Vol. 69, No. 4, 1980, pp. 531-538.

66. Anifantis, N. and Dimarogonas, A. D., "Identification of Peripheral Cracks in Cylindrical Shells," ASME Paper No. 83-WA/DE-14, 1983.
67. Dentsoras, A. J. and Dimarogonas, A. D., "Resonance Controlled Fatigue Crack Propagation in Cylindrical Shafts Under Combined Loading," ASME Paper No. 83-WA/DE-26, 1983.
68. Chamis, C., Sinclair, J. H. and Lark, R. F., "Dynamic Response of Damaged Angleplied Fiber Composites," Modern Developments in Composite Materials and Structures, ASME, 1979, pp. 31-51.
69. Tracy, J. J., Dimas, D. J. and Pardeon, G. C., "Advanced Composite Damage Detection Using Modal Analysis Techniques," Proceedings of the Second Modal Analysis Conference, 1984, pp. 665-670.
70. Crema, L. B., Castellani, A. and Peroni, I., "Modal Tests on Composite Material Structures: Application in Damage Detection," Proceedings of the Third Modal Analysis Conference, 1985, pp. 708-713.
71. Tsai, T., Yang, J. C. S. and Chen, R. Z., "Detection of Damages in Structures by the Cross Random Decrement Method," Proceedings of the Third International Modal Analysis Conference, 1985, pp. 691-700.
72. Akgun, M., Ju, F. D. and Paez, T. L., "Transmissibility as a Means to Diagnose Damage in Structures," Proceedings of the Third International Modal Analysis Conference, 1985, pp. 701-707.
73. "Modal Analysis Methods Applied to Structural Integrity Monitoring," Structural Dynamics Research Corporation, June 1985.
74. West, Walter, Fault Detection in Orbiter OV-101 Structure Related Structural Test Specimen, Loads and Structural Dynamics Branch Report, NASA/Johnson Space Center, September 1984.
75. West, Walter, "Single Point Random Modal Test Technology Application to Failure Detection," Shock and Vibration Bulletin, No. 52, Part 4, 1982, pp. 25-31.
76. West, Walter, "Illustration of the Use of Modal Assurance Criterion to Detect Structural Changes in an Orbiter Test Specimen," Proceedings of the Fourth International Modal Analysis Conference, 1986, pp. 1-6.

77. Haisty, B. S., West, W. and Mitchell, C., "Damage Detection in the Space Shuttle Orbiter Body Flap Test Specimen Using the Modal Assurance Criterion," Proceedings of the Advanced Composites Conference, 1985.
78. Yao, J. T. P., Toussi, S. and Sozen, M. A., "Damage Assessment From Dynamic Response Measurements," Proceedings of the Ninth United States National Congress of Applied Mechanics, 1982, pp. 315-322.
79. Yao, J. T. P., "Probabalistic Methods for the Evaluation of Seismic Damage in Existing Structures," Soil Dynamics and Earthquake Engineering, Vol. 1, No. 3, 1982.
80. Stephens, J. E. and Yao, J. T. P., "Survey of Available Structural Response Data for Damage Assessment," University of California Paper No. CE-STR-83-23, March 1984.
81. Adams, R. D. and Coppendale, J., "Measurement of the Elastic Moduli of Structural Adhesives by a Resonant Bar Technique," Journal of Mechanical Engineering Science, Vol. 18, No. 3, 1976, pp. 149-158.
82. Cempel, C., "Diagnostically Oriented Measures of Vibroacoustical Processes," Journal of Sound and Vibration, Vol. 73, No. 4, 1980, pp.547-561.
83. Cheng, J. S., Dunham, M. and Joanides, J., "Acoustic Test Response of the Space Shuttle Orbiter Bcdy Flap," Proceedings of the Second International Modal Analysis Conference, 1984, pp. 635-637.
84. Griesbach, T. J. and Ayers, D. J., "Opening and Extension of Circumferential Cracks in a Pipe Subject to Dynamic Loads," Nuclear Engineering Design, Vol. 57, No. 1, April 1980, pp. 141-152.
85. Koester, D. L., "The Dynamic Behavior of Center Crack Tension Specimens During High Frequency Material Fatigue Tests," Proceedings of the Third International Modal Analysis Conference, 1985, pp. 794-801.
86. Mutch, Gordon F. and Russell, Richard, "Troubleshooting In-Plant Equipment Using Combined Test and Analysis," Proceedings of the First International Modal Analysis Conference, 1982, pp. 252-258.
87. Sproat, W. H. and Rowe, W. J., "Ensuring Aircraft Structural Integrity Through Nondestructive Evaluation," Lockheed-Georgia Company.

88. Verdonck, E. and Snoeys, R., "Life Time Prediction Based on the Combined Use of Finite Element and Modal Analysis Data," Proceedings of the Second International Modal Analysis Conference, 1984, pp.572-579.

N89-21752

**SSME PROPULSION PERFORMANCE
RECONSTRUCTION TECHNIQUES**

Prepared by:	Enoch C. Temple, Ph.D.
Academic Rank:	Associate Professor
University and Department:	Alabama A&M University Department of Mathematics
NASA/MSFC Laboratory:	Propulsion
Division:	Propulsion Systems
Branch:	Performance Analysis
MSFC Colleague:	Klaus Gross
Date:	August 1988
Contract No.:	NGT-01-022-099 The University of Alabama

SSME PROPULSION PERFORMANCE RECONSTRUCTION TECHNIQUES

by

Enoch C. Temple

Associate Professor of Mathematics
Alabama A&M University
Normal, Alabama

ABSTRACT

In view of the complex flight operation of the Space Shuttle propulsion system together with an expected launch rate increase, the flight performance reconstruction process needs to be performed by automated computer programs. These programs must have the capability to quickly and reliably determine the true behavior of the various components of the propulsion system. For the flight reconstruction, measured values from the solid rocket motors, liquid engines, and trajectory are appraised through the Kalman filter technique to identify the most likely flight propulsion performance.

A more detailed data collection program for the single SSME engine captive test firing evaluation is scheduled for startup in September of 1988. Engine performance evaluation for the captive test firing requires a reconstruction process that is similar to the process that is used for the flight reconstruction. This paper describes analytical tools that may be used to reconstruct a propulsion system's true performance under flight and/or test conditions.

ACKNOWLEDGMENTS

I would like to thank NASA, MSFC, and ASEE for providing this excellent research opportunity. Thanks are also extended to the staff of the Propulsion Laboratory for the help willingly given during this ten-week period. Special thanks go to Klaus Gross, my NASA counterpart, for suggesting and discussing the topic addressed in this report. Further appreciation is extended to Judy Maples who typed this report.

1. INTRODUCTION

During the summer of 1987, Rogers Engineering and Associates (REA) submitted a technical report to Marshall Space Flight Center. This report summarized the results of the Propulsion Estimation Development Verification performed by REA under contract NAS8-36152. During that contract period, REA modified an existing program, which was developed under a previous contract, to include improved models of the Solid Rocket Booster, the Space Shuttle Main Engine (SSME) gain coefficient model, the vehicle trajectory using quaternions, and an improved Kalman filter based on the U-D factorized algorithm. In this report, this modified program will be identified by the symbols PROG1.

Under a current contract, REA has proposed to design, evaluate, and refine a model for a single SSME that does not include the influences of other Space Shuttle components. The refined SSME model is expected to permit the collection of high quality measurements and to provide an improved quality of SSME performance estimation. REA's single SSME model should be available in the Fall of 1988, and this report refers to that program by the symbols PROG2.

Propulsion performance estimation procedures applied in programs PROG1 and PROG2 are basically the same. Each requires the assumption that a correct model of the Space Shuttle's propulsion system exists and the model involves two main equations. They are an equation that models the dynamics of a state vector and a second equation that defines the relationship between an observation vector and the state vector. Also, the engine performance estimation consists of using collected data on the observation vector to obtain a corresponding estimate of the state vector. This state vector estimate is called an estimate or a reconstruction of the engine's performance.

The state vector estimate is based on the satisfaction of a statistical optimization criterion. Furthermore, data quality determination and the quality of the estimated state vector are judged through an application of statistics tools. Hence, in order for the Propulsion Laboratory to properly apply and correctly interpret the results of programs PROG1 and PROG2, an intuitive review of estimation procedures utilized by these programs is needed.

Therefore, the objectives of this paper are to:

1. Examine the assumptions and limitations of the Kalman filtering technique for the intended application to flight performance reconstruction of propulsion systems and single engine performance reconstruction based on static testing data.
2. Select specific topics from the area of statistics and probability for discussion and better understanding with EP55 personnel.

3. Identify some additional applications of contractor's programs.

Throughout this paper, underlined capital letters are used to denote vectors, capital letters denote matrices, and the identity matrix is denoted by the capital letter I . All vectors are of the column type and the transpose of any matrix or vector is denoted by using the letter T at the superscript position. The letter E denotes the expectation operator, $N(\underline{U}, \underline{\Sigma})$ denotes the multivariate normal probability distribution with mean vector \underline{U} and covariance matrix $\underline{\Sigma}$. The caret symbol " \wedge " written directly above a scalar or vector denotes a statistical estimator of that scalar or vector.

2. NOTATION

For any Space Shuttle flight or single SSME test, let \underline{Z}_t denote the observed values of vector \underline{Z} at time t . Each component of \underline{Z}_t represents a relevant measurable output of the Space Shuttle Propulsion System or a relevant navigation measurement. For example, the components of \underline{Z}_t may be oxygen pressurant flow, fuel volume flow, hydrogen pressurant flow, fuel flow pressure, fuel flow temperature, and chamber pressure whenever \underline{Z}_t represents the output of a single SSME static test. For an actual flight of the Space Shuttle, \underline{Z}_t will contain all of the aforementioned measurements for each of the three main engines plus additional components that represent navigational measurements and one measurement for each of the two Solid Rocket Motors (SRM). For the actual flight, Rogers (1987) lists 35 components for vector \underline{Z}_t and 71 components for state vector \underline{X}_t where \underline{X}_t is defined in the next paragraph. However, for the single engine static test setup, vectors \underline{Z}_t and \underline{X}_t have fewer components because the SRM's and navigational components become inactive.

Vector \underline{X}_t is a state vector of parameters to be estimated at time t . It is assumed that the observation vector \underline{Z}_t is a function of the state vector \underline{X}_t . That is,

$$\underline{Z}_t = \underline{h}(\underline{X}_t, t) + \underline{V}_t \quad (2.1)$$

where \underline{h} is some function and $\underline{V}_t \sim N(\underline{0}, \underline{R}_t)$. State vector \underline{X}_t is known to change with respect to time according to the equation

$$\dot{\underline{X}}_t = \underline{f}(\underline{X}_t, t) + \underline{W}_t \quad (2.2)$$

where \underline{f} is some function, $\underline{W}_t \sim N(\underline{O}, Q_t)$ and $E[\underline{W}_t \underline{V}_t^T] = 0$.

If it is assumed that \underline{f} and \underline{h} in equations (2.1) and (2.2) are linear, then numerical procedures allow us to transform these equations into the form

$$\underline{X}_k = \phi_{(k,k-1)} \underline{X}_{k-1} + \underline{W}_k \quad (2.3)$$

$$\underline{Z}_k = H_k \underline{X}_k + \underline{V}_k \quad (2.4)$$

where k represents discrete values of t , $\underline{W}_k \sim N(\underline{O}, Q_k)$, $\underline{V}_k \sim N(\underline{O}, R_k)$ and

$$E(\underline{W}_k \underline{V}_k^T) = E(\underline{W}_i \underline{W}_j^T) = E(\underline{V}_i \underline{V}_j^T) = 0, \quad \text{provided } i \neq j.$$

The symbol $\phi_{(k,k-1)}$ denotes the transition matrix that connects vector \underline{X}_k to \underline{X}_{k-1} .

After a Space Shuttle flight has taken place or after a static test, the post flight/test reconstruction procedure seeks to use the observed values of \underline{Z}_k and equations (2.3) and (2.4) to reconstruct (estimate) the vector \underline{X}_k so that the estimation error is minimized. The reader should see Gelb (1974) for a review of minimization procedures. The estimated value of \underline{X}_k is denoted by $\hat{\underline{X}}_k$.

3. THE KALMAN FILTER SETUP

Equations (2.3) and (2.4) along with assumed properties of vectors \underline{W}_k and \underline{V}_k allow for the development of the below Kalman filtering process. During the process, two types of estimators of \underline{X}_k are possible for each k . A notation used by Gelb (1974) allows the two estimators to be distinguished. That notation is

$\hat{\underline{X}}_k^{(-)} =$ the estimate of \underline{X}_k using all observations up to and including observation $(k-1)$.

$P_k^{(-)} =$ the estimate of the error covariance matrix for \underline{X}_k using all observations up to and including observation $(k-1)$.

$\hat{\underline{X}}_k^{(+)}$ = the estimate of \underline{X}_k using all observations up to and including observation k.

$P_k^{(+)}$ = the estimate of the error covariance matrix for \underline{X}_k using all observations up to and including observation k.

The values of $\hat{\underline{X}}_k^{(-)}$ and $P_k^{(-)}$ are obtained by using equation (2.3). The $\hat{\underline{X}}_k^{(-)}$ vector is often called the projected ahead value of \underline{X}_k and $P_k^{(-)}$ is the projected ahead variance. Vector $\hat{\underline{X}}_k^{(+)}$ is called the updated estimate of \underline{X}_k and P_k^{+} denotes the updated variance. Computation equations (3.1) through (3.5) describe the discrete Kalman filtering process.

$$\hat{\underline{X}}_k^{(-)} = \phi(k, k-1) \hat{\underline{X}}_{k-1}^{(+)} \quad (3.1)$$

$$P_k^{(-)} = \phi(k, k-1) P_{k-1}^{(+)} \phi^T(k, k-1) + Q_{k-1} \quad (3.2)$$

$$K_k = P_k^{(-)} H_k^T (H_k P_k^{(-)} H_k + R_k)^{-1} \quad (3.3)$$

$$\hat{\underline{X}}_k^{(+)} = \hat{\underline{X}}_k^{(-)} + K_k (\underline{Z}_k - H_k \hat{\underline{X}}_k^{(-)}) \quad (3.4)$$

$$P_k^{+} = (I - K_k H_k) P_k^{(-)} \quad (3.5)$$

To start the computation procedure, initial estimates $\hat{\underline{X}}_0^{(-)}$ and $P_0^{(-)}$ are needed and are usually determined by the users of equations (3.1) through (3.5).

The linear system described in equations (2.3) and (2.4) is essential to the development of the discrete Kalman filtering process given in equations (3.1) through (3.5). Therefore, when functions f and h are nonlinear, as they are in the case of the flight/test reconstruction model, vector \underline{X}_k has to be estimated through a linearization process. Two different linearization procedures are described in the next paragraph. The first procedure is called the linearized filter and the other is known as the extended Kalman filter.

Let \underline{X}_k^* be some known vector such that $\dot{\underline{X}}_k^* = \underline{f}(\underline{X}_k^*, k)$. Often times \underline{X}_k^* is called a reference solution or a reference nominal trajectory. That is, \underline{X}_k^* is a known solution to equation (2.2) where the influence of \underline{W}_t is not considered. If a discrete solution to equation (2.2) is \underline{X}_k where $\underline{X}_k = \underline{X}_k^* + \Delta \underline{X}$, then the linearization process may be used to estimate $\Delta \underline{X}$. Here, we see that for any $\Delta \underline{X}$ and for any k value of t , a first degree Taylor series approximation of functions (2.1) and (2.2) may be determined by

$$\dot{\underline{X}}_k + \Delta \dot{\underline{X}} \approx \underline{f}(\underline{X}_k^*, k) + \left[\frac{\partial \underline{f}}{\partial \underline{X}} \right]_{\underline{X}=\underline{X}_k^*} \Delta \underline{X} + \underline{W}_k \quad (3.6)$$

and

$$\underline{Z}_k \approx \underline{h}(\underline{X}_k^*, k) + \left[\frac{\partial \underline{h}}{\partial \underline{X}} \right]_{\underline{X}=\underline{X}_k^*} \Delta \underline{X} + \underline{V}_k \quad (3.7)$$

Since $\dot{\underline{X}}_k = \underline{f}(\underline{X}_k, k)$, equations (3.6) and (3.7) reduce to

$$\Delta \dot{\underline{X}} = \left[\frac{\partial \underline{f}}{\partial \underline{X}} \right]_{\underline{X}=\underline{X}_k^*} \Delta \underline{X} + \underline{W}_k \quad (3.8)$$

and

$$\underline{Z}_k - \underline{h}(\underline{X}_k^*, k) = \left[\frac{\partial \underline{h}}{\partial \underline{X}} \right]_{\underline{X}=\underline{X}_k^*} \Delta \underline{X} + \underline{V}_k \quad (3.9)$$

Equation (3.8) is called the linearized dynamics equation and (3.9) is the linearized measurement equation. These two equations are linear and are equivalent to the linear equations (2.3) and (2.4). Hence, for each discrete time k , $\Delta \underline{X}_k$ can be estimated and error covariance matrices can be determined. The state vector estimate at time k is then given by

$$\hat{\underline{X}}_k^+ = \underline{X}_k^* + \Delta \hat{\underline{X}}_k^+ \quad (3.10)$$

where $\hat{\underline{X}}_k^+$ is computed by applying the Kalman filtering process to equations (3.8) and (3.9). When each \underline{X}_{k+1}^* ($k = 0, 1, \dots, T$) is known prior to the beginning of the application of the Kalman filtering process to equations (3.8) and (3.9), the estimate in equation (3.10) is called the linearized Kalman estimate. When each \underline{X}_{k+1}^* is determined from the previous estimate of \underline{X}_k , i.e., $\underline{X}_{k+1}^* = \hat{\underline{X}}_k^+$, the estimate in equation (3.10) is called the extended Kalman filter estimate.

4. FAILURE DETECTION

The current replacement cost of a single SSME is about \$50 million. This cost, combined with analyses of data from static test failures and shutdowns, suggest that there is a need for a more advanced state test failure detection system for the SSME. A 1985 report by Taniguchi emphasizes the importance of SSME failure detection improvements and identified some possible analytic designs for a failure detection system. Each of Taniguchi's failure detection designs assumed that a correct model of the single SSME exists. Each design also required the application of a statistics test as a decision tool. Therefore, a natural extension of REA's SSME static test model would be into the area of failure detection during static testing. REA's static test model is already undergoing refinements and is expected to be available to Marshall Space Flight Center by the Fall of 1988. Also, according to Taniguchi (1985), data is available on 1200 static test firings of a single SSME. A brief overview of a failure detection setup is provided in the next paragraph.

Assume that equations (2.3) and (2.4) are modified to yield

$$\underline{X}_k = \Phi(k, k-1) \underline{X}_{k-1} + A_1(k, k-1) \underline{b}_k + \underline{W}_k \quad (4.1)$$

$$\underline{Z}_k = H_k \underline{X}_k + A_2(k) \underline{b}_k + \underline{V}_k \quad (5.2)$$

where vectors \underline{b}_k are called biased vectors. If biases are regarded as additional states with dynamics such that

$$\underline{b}_{k+1} = \underline{b}_k$$

then equations (4.1) and (4.2) may be transformed into the form

$$\underline{Y}_k = \begin{bmatrix} \Phi(k, k-1) & A_1(k, k-1) \\ 0 & I \end{bmatrix} \underline{Y}_{k-1} + \begin{bmatrix} I \\ 0 \end{bmatrix} \underline{W}_k \quad (4.3)$$

and

$$\underline{Z}_k = [H_k \quad A_2(k)] \underline{Y}_k + \underline{V}_k \quad (4.4)$$

where

$$\underline{Y}_k = \begin{bmatrix} \underline{X}_k \\ \underline{b}_k \end{bmatrix} .$$

If Kalman filtering is applied to equations (4.3) and (4.4), the bias vector \underline{b}_k is estimated along with the other components of the state vector.

If no failure has occurred at time k , then it is intuitively reasoned that $\underline{b}_k = \underline{0}$. Of course, the judgement as to whether \underline{b}_k is really zero is determined by a multivariate statistical test.

For additional details on bias vector estimation, the reader should consult Friedland (1983).

5. CONCLUSION

The assumptions and limitations of the Kalman filtering process have been discussed with NASA personnel. These discussions included relevant topics from the area of statistics. In particular, it was pointed out in this paper that equations (2.3) and (2.4) are the essential ingredients for the discrete variable reconstruction. Hence, if functions \underline{f} and \underline{h} of equations (2.1) and (2.2) are nonlinear, \underline{X}_k can be estimated by linearizing \underline{f} and \underline{h} about some nominal vector solution \underline{X}_k^* .

This report also points out that REA's SSME model may be modified to form a failure detection procedure. The reader should be reminded that the intuitive failure detection approach outlined from equations (4.3) and (4.4) is based on an assumption that \underline{f} and \underline{h} of equations (2.1) and (2.2) are linear functions. Since functions \underline{f} and \underline{h} are generally nonlinear for SSME applications, it is worthwhile to investigate the effectiveness of bias estimation techniques as an SSME static test failure detection device.

REFERENCES

- Gelb, A. (Editor, 1974), Applied Optimal Estimation, Cambridge: MIT Press.
- Friedland, B. (1983), "Separated-Bias Estimation and Some Applications," in Control and Dynamic Systems, Advances in Theory and Application, Edited by C. T. Leondes, New York: Academic Press.
- Rogers, R. M. (1987), Space Shuttle Propulsion Estimation Development Verification, Final report for Contract NAS8-36152.
- Taniguchi, M. H. (Editor, 1985), Failure Control Techniques for the SSME. Final Report, NAS8-36305, Rocketdyne Division.

N89-21753

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

DYNAMIC FATIGUE TESTING OF ZERODUR GLASS-CERAMIC

Prepared by:	Dennis S. Tucker
Academic Rank:	Assistant Professor
University and Department:	Georgia Tech
NASA/MSFC:	
Laboratory:	Materials and Processes
Division:	Non-Metallic Materials
Branch:	Ceramics and Coatings
NASA Colleague:	Ron L. Nichols
Contract No.:	NGT-01-002-009 The University of Alabama

ABSTRACT

The inherent brittleness of glass invariably leads to a large variability in strength data and a time dependence in strength. Loading rate plays a large role in strength values. Glass is found to be weaker when supporting loads over long periods of time as compared to glass which undergoes rapid loading. These properties complicate the structural design allowables for the utilization of glass components in an application such as the AXAF.

This report describes the test methodology to obtain parameters which can be used to predict the reliability and lifetimes of Zerodur glass-ceramic which is to be used for the mirrors in the AXAF.

ACKNOWLEDGEMENT

The author wishes to acknowledge the NASA/ASEE Summer Faculty Fellowship Program along with Mike Freeman, the University of Alabama director.

To NASA counterpart Ronald L. Nichols of the Ceramics and Coatings Branch, Non-Metallic Materials Division, gratitude is offered.

List of Figures

Figure No.

Caption

- | | |
|---|--|
| 1 | Double Ring Fixture |
| 2 | Design Diagram for Polished
ULE Glass |
| 3 | Design Diagram for Unpolished
ULE Glass |

INTRODUCTION

The aim of this study is to provide the necessary data to predict the reliability and lifetime of Zerodur Glass-Ceramic which will be utilized as the material for the grazing incidence mirrors in the Advanced X-ray Astrophysics Facility (AXAF).

Zerodur is a alumino-silica glass-ceramic containing 70 to 78% by weight of high-temperature quartz as the crystalline material. The mean crystal size of the crystals is in the region of 50 to 55 NM. The crystalline phase has a negative expansion coefficient, while the vitreous phase has a positive expansion coefficient. This practically balances out linear thermal expansion leading to long term stability even with temperature fluctuations. Zerodur, is therefore an ideal candidate for AXAFS and other applications such as mirror substrates mounts where changes in the material due to temperature variations could impair the quality of observations. The composition and mechanical, thermal and optical properties have been reported by Schott Glass Technologies, producers of Zerodur. However, the question of subcritical crack growth in this material has not been explored in detail.

The design of the stresses in the AXAF mirrors is dictated by the ultimate strength of the material and a nominal value of 1450 psi (MOR) is recommended by the manufacturer. This is a minimum value or worst case value. In glassy materials, the modulus of rupture depends upon primarily the surface condition of the material. In other words, a surface with a large number of flaws will have a lower strength than one with few or no flaws. The reduction in strength is caused by surface flaws which concentrate applied stresses leading to failure at much lower than theoretical applied loads. These flaws can be introduced by machining or normal handling. Thus one method to increase strength is to polish the surface thereby removing surface flaws. Schott Glass reports an MOR of approximately 11,000 psi with a failure probability of 0.5% when the surface is polished with 600 grit polishing paper. Even higher strengths could be expected with a finer polishing medium. Even if one were to accept a value in the range of 11,000 psi one must consider the effects of delayed failure under constant applied loads. Under these loads, even if small, failure can occur after a given period of time has passed. This has been attributed to the mechanism of stress corrosion whereby water molecules from a humid environment enhance crack growth. Cracks continue to grow over a period of time until catastrophic failure occurs.

Therefore one needs a reliable method of predicting life times under applied loads. The fatigue parameters for making lifetime predictions must be obtained under the service environment, if at all possible and can be obtained from one of three types of experiments: crack velocity, static fatigue and dynamic fatigue.

Crack velocity can be measured directly as a function of the stress intensity factor on specimens specifically designed for fracture mechanics experiments.¹ These specimens contain a macroscopic crack which allows accurate measurement of crack velocity and stress intensity factor. For purposes of failure prediction crack velocity data is not as reliable as that from static or dynamic fatigue strength experiments because data from large, preformed cracks is not necessarily relevant to the propagation of microscopic cracks present in brittle materials.²

Static fatigue data is generally obtained by measuring the time to failure of a large number of samples at several constant applied stresses.³

Dynamic fatigue is generated by measuring fracture strength as a function of stress rate.³ For this study dynamic fatigue was chosen to develop the time-to-failure equation necessary to a design diagram. This concept will be developed fully in the Technical Discussion section.

Due to problems in obtaining samples in time to perform the actual experiments no data will be presented in the report. Instead attention will be paid to theoretical development of the design diagram and the proposed experimental procedure to obtain the necessary data for the diagram.

Technical Discussion

Initial Strength

Modulus of Rupture (MOR) will be determined using the double ring bending strength test method.⁴ Figure 1 is a graphical representation of this technique. In this case a disc shaped specimen is loaded between two concentric rings. For limited forces, a tensile stress field will be set up in the central region of the convexly bent specimen surface. Outside the load ring the radial and tangential stresses in the specimen decrease toward the edge, so that the possibility of fracture there is small. By increasing the force the tensile stress in the specimen center is increased at a constant rate until fracture occurs, with the expectation that the fracture is initiated in the region of the surface subjected to the maximum stress, underneath the load ring.

The major advantage of this technique, as compared to techniques such as 3 and 4 point bending, is the elimination of edge effects.

To calculate the rate of increase of the bending stress in circular specimens the following is applicable:

$$\frac{\Delta \sigma_b}{\Delta t} = \frac{1.08}{S^2} \cdot \frac{\Delta F}{\Delta t} \quad (1)$$

Where:

ΔF = Increase of test force, measured in the time interval near the fracture initiation

ΔS = Specimen thickness

1.08 = Numerical constant related to the load ring and specimen support ring diameters and Poisson's Ratio

The bending strength (σ_{bB}) belonging to the maximum force (F_{max}) from equation (1) is:

$$\sigma_{bB} = 1.08 \frac{F_{max}}{S^2} \quad (2)$$

Where:

σ_{bB} = bending strength

F_{max} = greatest force

S = specimen thickness

Weibull Statistics

In order to determine the distributions of strength Weibull plots will be utilized. Fracture of brittle materials can be regarded as a statistical process, since the material has a distribution of flaw sizes. Weibull⁵ described the mathematics for this process in terms of probability functions. His equations can take forms:

$$F = 1 - \exp \left(-\frac{\sigma}{\sigma_0} \right)^m \quad (3)$$

or

$$F = 1 - \exp \left(-\frac{\sigma - \sigma_u}{\sigma_0} \right)^m \quad (4)$$

where:

F = probability that the material will fail at stress

σ_u, σ_0 = constants which describe the distribution

m = constant (Weibull modulus)

Equation (3) is the two-parameter model and represents an unbounded distribution. Equation (4) is the three-parameter model and represents a lower bounded distribution.

A number of dynamic fatigue studies of glass and glass-ceramics have shown the two-parameters model to be acceptable in describing the flaw population and it's relation to material failure.^{6, 7, 8} Therefore, in this study equation (3) will be utilized.

One can rearrange equation (3) to obtain a more workable equation. Starting with equation (3)

$$F = 1 - \exp \left(-\frac{\sigma}{\sigma_0} \right)^m \quad (3)$$

$$-1 + F = -\exp \left(-\frac{\sigma}{\sigma_0} \right)^m$$

$$1 - F = \exp \left(-\frac{\sigma}{\sigma_0} \right)^m$$

$$(1 - F)^{-1} = \exp \left(\frac{\sigma}{\sigma_0} \right)^m$$

or

$$\frac{1}{1 - F} = \exp \left(\frac{\sigma}{\sigma_0} \right)^m \quad (5)$$

Taking \ln of both sides of (5)

$$\ln \left(\frac{1}{1-F} \right) = \left(\frac{\sigma}{\sigma_0} \right)^m \quad (6)$$

and taking ln again;

$$\ln \left(\ln \frac{1}{1-F} \right) = m/N \ln \sigma - m/N \ln \sigma_0 \quad (7)$$

Where:

σ = Failure stress (MOR)

And:

$$F = \frac{m - 0.5}{N}$$

Where:

m = Rank of sample in terms of strength

N = Number of samples tested

Thus a plot of $\ln \left(\ln \frac{1}{1-F} \right)$ versus $\ln \sigma$ will yield a slope of m . The slope indicates the distribution of strengths. As the slope (m) increases the distribution narrows, i.e. there is less scatter in the results.

Lifetime Prediction Equation

Weibull statistics can be combined with data on static or dynamic fatigue to yield predictions of the lifetime of brittle materials under various environmental conditions.³ The derivation of this equation combines elements of fracture mechanics theory and Weibull statistics.

The rate of crack growth in a material can be expressed as a power law function:

$$v = \frac{da}{dt} = AK_I \quad (8)$$

Where:

a = crack length

K_I = stress intensity factor

A, N = constant for a given material and environmental

The stress intensity factor can be expressed as:

$$K_I = \sigma Y \sqrt{a} \quad (9)$$

Where:

a = crack length

σ = applied stress

Y = crack geometry parameter which is constant for a given material

Combining equations (8) and (9) to eliminate K_I yields:

$$\frac{da}{dt} = A \sigma^N Y^N a^{N/2} \quad (10)$$

For the case of constant applied stress (σ_a):

$$\frac{da}{dt} = A \sigma_a^N Y^N a^{N/2}$$

or

$$a^{-N/2} da = A \sigma_a^N Y^N dt \quad (11)$$

To obtain the material lifetime, this expression is integrated over the initial-to-final crack length and from the time the stress was applied ($t=0$) to the time of fracture ($t = t_f$).

$$\int_{a_0}^{a_f} a^{-N/2} da = \int_0^{t_f} A \sigma_a^N Y^N dt \quad (12)$$

$$\frac{a^{-N/2+1}}{-N/2+1} \Big|_{a_0}^{a_f} = A \sigma_a^N Y^N t_f \Big|_0^{t_f}$$

or

$$\frac{2a^{-(N-2/2)}}{N-2} \Big|_{a_0}^{a_f} = A \sigma_a^N Y^N t_f \Big|_0^{t_f}$$

which yields:

$$\frac{2}{N-2} \left(a_f^{-(N-2/2)} - a_0^{-(N-2/2)} \right) = A \sigma_a^N Y^N t_f \quad (13)$$

Rearranging equation (9):

$$a = \left(\frac{K_{If}}{\sigma_f Y} \right)^2$$

and substituting into equation (13):

$$\frac{2}{N-2} \left[\left(\frac{K_{If}}{\sigma_f Y} \right)^{-(N-2)} - \left(\frac{K_{I0}}{\sigma_0 Y} \right)^{-(N-2)} \right] = A \sigma_a^N Y^N t_f \quad (14)$$

The stress intensity factor when fracture occurs, K_{IC} , and the fracture stress σ_f is the stress at which fracture would occur in an inert environment unaffected by subcritical crack growth (σ_i)

Substituting into equation (14):

$$\frac{2}{N-2} \left[\left(\frac{K_{IC}}{\sigma_i Y} \right)^{-(N-2)} - \left(\frac{K_{IO}}{\sigma_o Y} \right)^{-(N-2)} \right] = A \sigma_a^N Y^N t_f \quad (15)$$

Generally $K_{IC}^{-(N-2)} \ll K_{IO}^{-(N-2)}$

Therefore one can write:

$$\frac{2}{N-2} \left(\frac{\sigma_i Y}{K_{IC}} \right)^{N-2} = A \sigma_a^N Y^N t_f \quad (16)$$

Solving for t_f :

$$t_f = \frac{2}{N-2} \frac{\sigma_i^{N-2} Y^{N-2}}{A \sigma_a^N K_{IC}^{N-2} Y^N} = \frac{2}{N-2} \frac{\sigma_i^{N-2}}{A \sigma_a^N K_{IC}^{N-2} Y^2} = B \sigma_i^{N-2} \sigma_a^{-N} \quad (17)$$

Where:

$$B = \frac{2}{A Y^2 K_{IC}^{N-2} (N-2)}$$

Taking \ln of both sides of equation (17) (17)

$$\ln t_f = \ln B + (N-2) \ln \sigma_i - N \ln \sigma_a \quad (18)$$

Rearrangement of Equation (7) yields:

$$\ln \left(\ln \frac{1}{1-F} \right) + m \ln \sigma_o = m \ln \sigma$$

or

$$\ln \sigma = \frac{1}{m} \left[\ln \left(\ln \frac{1}{1-F} \right) + m \ln \sigma_o \right]$$

(19)

In an environment where no subcritical flaw growth occurs prior to fracture one can write equation (19) in terms of inert strength (σ_i) i.e.

$$\ln \sigma_i = \frac{1}{m_i} \left[\ln \left(\ln \frac{1}{1-F} \right) + m_i \ln \sigma_i \right] \quad (20)$$

Substitution of equation (20) into equation (18) for $\ln \sigma_i$ yields:

$$\ln t_f = \ln B + \frac{N-2}{m_i} \left[\ln \left(\ln \frac{1}{1-F} \right) + m_i \ln \sigma_i \right] - N \ln \sigma_a \quad (21)$$

Where:

t_f represents the time required for a flaw to grow from an initial subcritical size to dimensions critical for catastrophic crack propagation leading to material failure.

And:

B and N are constants which characterize the subcritical flaw growth.

Determination of B and N

Evans 10 derived an expression relating the strength of a material (σ) to the stressing rate ($\dot{\sigma}$) at constant temperature for the case of subcritical flaw growth as follows:

$$\frac{d\sigma}{dt} = \dot{\sigma} \quad (22)$$

or,

$$d\sigma = \dot{\sigma} dt \quad (23)$$

Dividing by da ;

$$\frac{d\sigma}{da} = \dot{\sigma} \frac{dt}{da} \quad (24)$$

Since $\frac{da}{dt} = V$ (crack velocity):

$$\frac{d\sigma}{da} = \frac{\dot{\sigma}}{V} \quad (25)$$

Now, we know that,

$$V = AK_I$$

Therefore we can write:

$$\frac{d\sigma}{AK_I^N} = \dot{\sigma} da \quad (26)$$

We also know:

$$K_I = Y \sigma \sqrt{a}$$

or

$$K_I^N = Y^N \sigma^N a^{N/2}$$

Substituting this expression into equation (26) gives:

$$d\sigma = \dot{\sigma} da / A \sigma^N a^{N/2} Y^N$$

or

$$\sigma^N d\sigma = \dot{\sigma} da / A Y^N a^{N/2} \quad (27)$$

Integrating over the extent of crack growth from an initial stress of zero to a final failure stress or obtains:

$$\int_0^{\sigma_f} \sigma^N d\sigma = \frac{\dot{\sigma}}{A Y^N} \int_{a_i}^{a_c} a^{-N/2} da \quad (28)$$

which yields:

$$\frac{\sigma_f^{N+1}}{N+1} = \frac{2\dot{\sigma}}{A Y^N (2-N)} \left[a_c^{(2-N)/2} - a_i^{(2-N)/2} \right] \quad (29)$$

Since, $a_c^{(2-N)/2} \gg a_i^{(2-N)/2}$

the former can be dropped leading to:

$$\frac{\sigma_f^{N+1}}{N+1} = \frac{2\dot{\sigma}(-a_i^{(2-N)/2})}{A Y^N (2-N)} \quad (30)$$

The initial flaw size can be taken as:

$$a_i^* = K_{IC} / Y \sigma_{ic}$$

or

$$a_i = \frac{K_{IC}^2}{Y^2 \sigma_{ic}^2} \quad (31)$$

Substituting equation (31) into (30) for a_i yields:

$$\frac{\sigma_f^{N+1}}{N+1} = \frac{2(N+1)\dot{\sigma} \sigma_{ic}^{N-2}}{A Y^2 (N-2) K_{IC}^{N-2}} \quad (32)$$

Can rewrite equation (32) as:

$$\nabla_f^{N+1} = B (N + 1) \nabla_{ic} \dot{V} \quad (33)$$

where:

$$B = \frac{2}{AY^2(N-2)K_{IC}^{N-2}}$$

Taking ln of both sides of equation (33):

$$\begin{aligned} (N + 1) \ln \nabla_f &= \ln [B (N + 1) + (N - 2) \nabla_{ic}] + \ln \dot{V} \\ \text{or} \\ \ln \nabla_f &= \frac{\ln [B (N + 1) + (N - 2) \nabla_{ic}] + \ln \dot{V}}{N + 1} \end{aligned} \quad (34)$$

Rewriting Equation (34):

$$\ln \nabla_f = \left(\frac{\ln [B (N + 1)] + (N - 2) \ln \nabla_{ic}}{N + 1} \right) + \frac{\ln \dot{V}}{N + 1} \quad (35)$$

or

$$\ln \nabla_f = a_0 + a_1 \ln \dot{V} \quad (36)$$

where:

$$\begin{aligned} a_0 &= \frac{\ln [B (N + 1)] + (N - 2) \ln \nabla_{ic}}{N + 1} \\ a_1 &= \frac{1}{N + 1} \end{aligned}$$

Thus, to obtain values for B and N, one fits fracture stress and stress rate data for samples tested in humid environment to a straight line given by equation (36), that is a plot $\ln \sigma_f$ versus $\ln \dot{\sigma}$. To ensure that the values of B and N are reliable one needs to test over a wide range of stress rates (i.e. over several orders of magnitude). The inert strength (σ_i) should be obtained at high enough stress rates to insure that crack growth is independent of environmental (i.e. no stress corrosion at the crack tip is occurring).

Design Diagrams

Once the values of the constants B and N are determined, the time-to-failure for various values of failure probability (F) and applied stress (σ_a) can be computed. Figure 2 shows an example of a design diagram for polished ULE glass at F = 0.001, 0.01 and 0.20. Comparing this to a diagram for unpolished ULE glass (Figure 3) one can note decrease in strength at a given time period (Ex. 10 yrs.) as compared to polished glass. This infers that at a given probability of failure the polished glass should withstand an applied stress for a longer period of time. For examples at F = 0.01 and an applied stress of 24×10^6 MPa the polished material will fail in one hour whereas the unpolished fails in one second.

Experimental Approach

The investigation will be carried out in three phases.

Phase I- Residual Stress Investigation

In order to determine the effects of residual stresses on the modulus of rupture of Zerodur, 10 samples with ground surface finish will be annealed and compared with 10 duplicate unannealed samples using the double ring bend test method. All tests will be performed at a stress rate of 290 psi/sec, which is in accordance with test standard DIN 52292.

Phase II - Generation of Design Diagram

Weibull plots [$\ln (\ln \frac{1}{1-F})$ vs. $\ln \sigma_f$] will be made in order to determine the strength distributions for three types of sample preparation (AS cut, ground, and ground with acid etch). The etching technique is to be supplied by Schott Glass Co. MOR data will be obtained at four different stress rates, 3.5×10^{-3} MPa/s, 2.7×10^{-2} MPa/s, 2 MPa/s and 200 MPa/s. Thirty five samples will be tested. The final stress at each stress rate will be used to determine the inert strength. In this instance samples will be heated at 150°C for 24 hours in a dry N₂ environment and put in a vacuum dessicator before testing.

Each sample will be exposed to flowing dry nitrogen for fifteen minutes before testing, in a plexiglass chamber fitted to the testing machine. To insure that this stress rate leads to crack growth independent of environment, ten samples will be placed in distilled water for 24 hours and then tested at 200 MPa/s. The sample to be tested at the three lower stress rates will be stored at ambient conditions ($T = 25^{\circ}\text{C}$, R.H. = 50%) for at least one week before testing.

Using the time-to-failure equation developed in this paper, a time-to-failure diagram will be developed for each sample preparation. That is a separate diagram for the as cut, ground, and ground and acid etched samples.

Phase III- Verification of Diagrams

In order to verify the three time-to-failure diagrams, applied stress values which intersect the $F = 0.005$ probability curve corresponding to time-to-failures of 1 minute, 1 hour, and 1 day will be selected and used to test 10 specimens each using the double ring test method.

Bibliography

1. S. M. Wiederhorn, "Subcritical Crack Growth in Ceramics," pp 613-46 in Fracture Mechanics of Ceramics, Vol. 2, ed. by R. C. Brandt, D. P. H. Hasselman, and F. F. Lange, Plenum Press, New York (1974).
2. J. E. Ritter, Jr., "Engineering Design and Fatigue Failure of Brittle Materials," pp. 667-686 in Fracture Mechanics of Ceramics, Vol. 4, Ed. by R. C. Brandt, D. P. H. Hasselman, and F. F. Lange, Plenum Press, New York (1978).
3. J. E. Ritter, Jr. and C. L. Sherburne, "Dynamic and Static Fatigue of Silicate Glasses," J. of Am. Ceram. Soc., 54 601-05, (1971).
4. "Testing of Glass and Glass Ceramics; Determination of Bending Strengths," German Standard DK 66.151: 620.174, DIN 52292 Part 1 April 1984.
5. W. Weibull, "A Statistical Theory of the Strength of Materials," Roy. Swed. Acad. of Eng. Sci. Proc. 151, 1-45 (1939).
6. R. J. Charles. "Dynamic Fatigue of Glass," J. of Appl. Phys. 29 [12] 1657-62, (1958).
7. K. K. Smyth and M. B. Magida, "Dynamic Fatigue of a Machinable Glass-Ceramic," J. Am. Ceram. Soc. 66 [7], 500-505 (1983).
8. D. C. Cramner and D. J. Speece, "Dynamic Fatigue of Ultralow-Expansion Glass for Space Mirrors," Report SD-TR-86-49, Space Div., Airforce Systems Command, Aug. 1986.
9. S. M. Wiederhorn, Fracture 1977 Ed. D. M. R. Taplin, Vol.3, pp. 893-901, Waterloo Univ. Press, Waterloo, Can. (1977).
10. Evans, A. G., Int. J. Fract. 10 251-259 (1974).

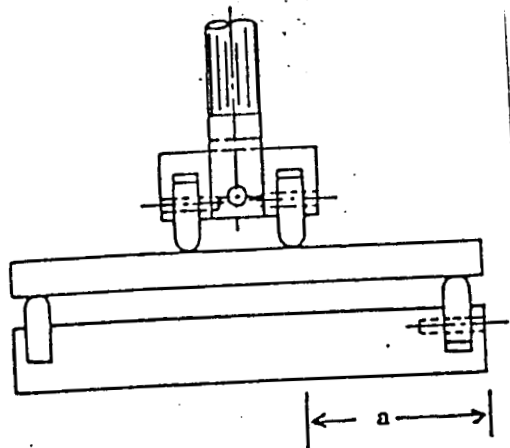


Figure 1

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

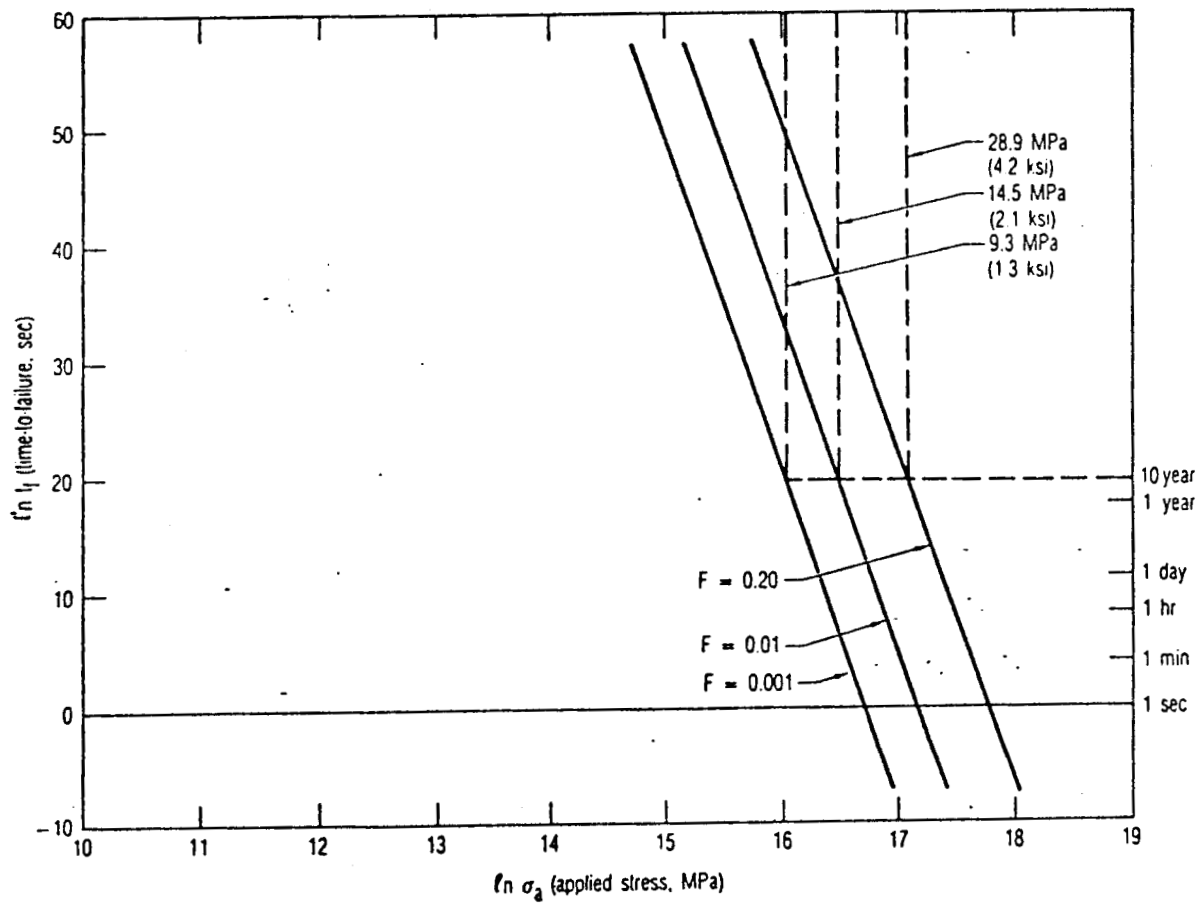


Figure 2

ORIGINAL PAGE IS
OF POOR QUALITY

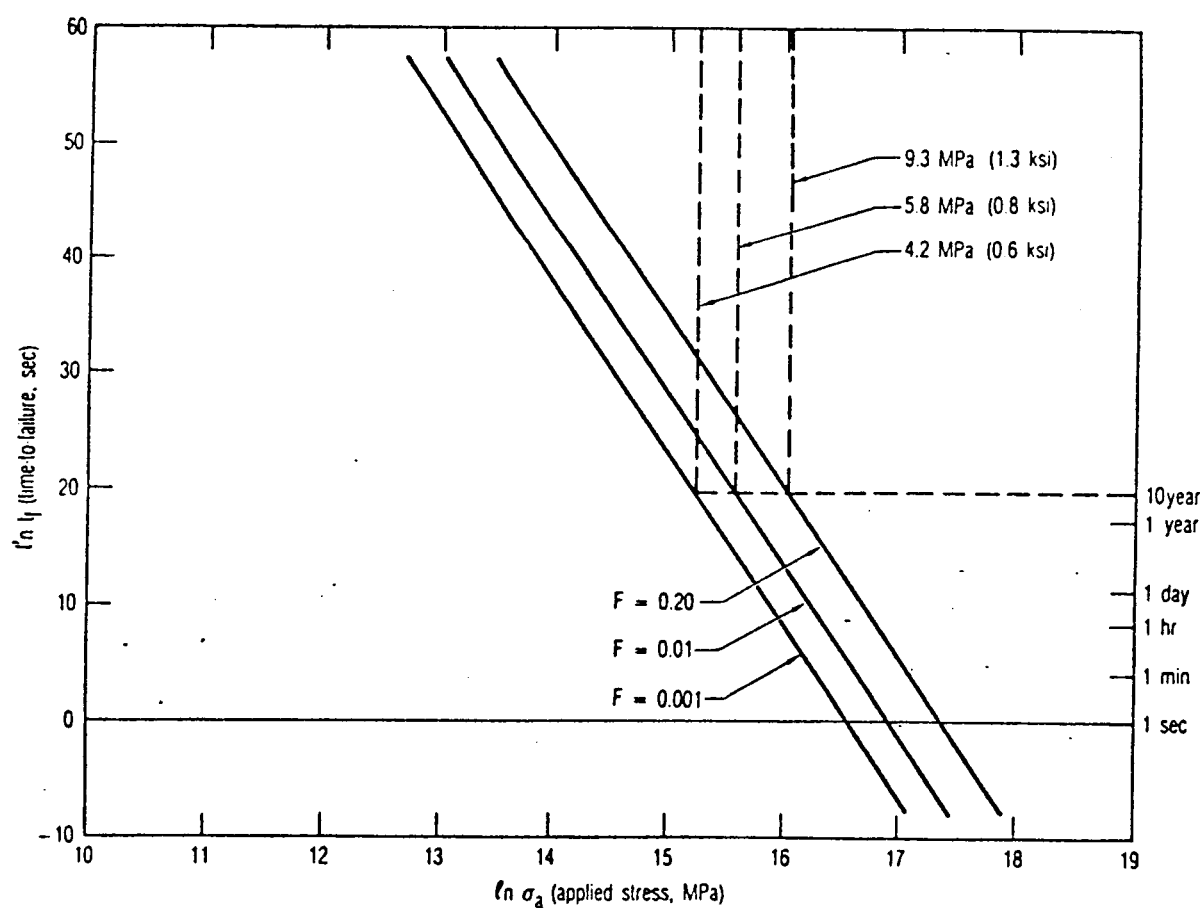


Figure 3

N89-21754

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

SOFTWARE SYSTEM SAFETY

Prepared by:	James G. Uber
Academic Rank:	Assistant Professor
University and Department:	The University of Alabama in Huntsville; Department of Mechanical Engineering
NASA/MSFC: Laboratory:	Safety, Reliability, Maintainability and Quality Assurance
Division: Branch:	Systems Safety Engineering Hazard Analysis
MSFC Colleague:	Dewey B. Channell
Date:	August 26, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

SOFTWARE SYSTEM SAFETY

by

James G. Uber
Assistant Professor of Civil Engineering
The University of Alabama in Huntsville
Huntsville, Alabama

Abstract

Software itself is not hazardous, but since software and hardware share common interfaces there is an opportunity for software to create hazards. Further, these software systems are complex, and proven methods for the design, analysis, and measurement of software safety are not yet available. This report reviews some past software failures, future NASA software trends, software engineering methods, and tools and techniques for various software safety analyses. Recommendations to NASA SRM&QA are made based on this review.

ACKNOWLEDGEMENTS

I wish to thank the many engineers in the SRM&QA Systems Safety Division who made my stay enjoyable and who shared their thoughts and concerns about software and systems safety. Also, it was Mike Freeman and Gerald Karr who introduced me to members of the SRM&QA organization and thus made this work possible. The financial assistance provided by the NASA/ASEE summer faculty fellowship program is acknowledged.

1. INTRODUCTION

Complex software systems are often required for real-time control and monitoring of aerospace systems hardware. The NASA Space Shuttle software, for example, consists of some ten million lines of computer code that control and monitor hardware on the ground and in the air; the Shuttle could not fly without it. Since software and hardware share common interfaces, there are opportunities for software to create hazards. This linkage has historically been ignored, however, in formal analyses of system safety.

This report examines some approaches for identifying and eliminating or controlling software hazards. In order to appreciate the software safety problem, however, it is necessary to understand something about software development practices, the characteristics of software and software errors, and "the reasons why things are the way they are." Accordingly, much of the report text is devoted to these preliminary issues in preparation for a later discussion of software safety and software safety analysis methods.

Section three provides the motivation for continuing on into the rest of the report. This motivation is provided by discussions of software development "horror stories," future trends in NASA software control systems development, and the value of implementing software safety programs. The next section discusses the current state of software engineering, some reasons for the current situation, and some modern approaches to software development. Section five considers the software safety problem and some solutions, and section six presents recommendations for addressing present and future software safety problems.

2. Objectives

The objectives of this study are to explore the breadth and depth of the software system safety problem, and to recommend SRM&QA actions that can reduce the present and future hazards posed by software.

3. MOTIVATION

3.1. Statistics and Failures

Does the Federal Government get what it pays for? Perhaps not, at least when it comes to software development. A decade-old study of the software costs associated with nine federal projects depicts software procurement as a monetary black hole [Neumann, 1985]. Of software expenditures totalling \$6.8 million, 95% (\$6.5 million) was wasted on software that either was delivered but never used (47%), paid for but never delivered (29%), or abandoned or reworked (19%). Less than 2% of the software expenditures yielded software that was used as delivered. These statistics reflect the inherent difficulty of large-scale software management and development, although improvements in software engineering practice may produce more impressive results if a similar study were conducted today.

Problems with real-time control software may be subtle and remain innocuous until the worst possible moment. These problems can also survive supposedly rigorous testing procedures. A good example of such a software problem is the "bug" that delayed the first shuttle orbital flight on April 10, 1981. The bug was discovered when, twenty minutes prior to scheduled launch, the backup flight control system (BFS) failed to initialize when commanded. Curiously, it turned out that the bug was not in the BFS, but was in the primary avionics software system (PASS), and had gone undetected for over a year. John Garman, who at the time was Deputy Chief of the Spacecraft Software Division at JSC, gives a detailed account of the discovery and correction of the bug [Garman, 1981]; some of the details of this very intricate and subtle problem are given below.

There are five general purpose computers (GPC's) onboard the Shuttle. These five independent machines combined with the PASS and BFS provide two different types of redundancy to guard against two different types of failures. The PASS executes on four of the GPC's and provides for full operational capability after the failure of one GPC and a safe return capability after the failure of a second. Note, however, that this redundancy provides protection only against hardware failures. Potential hazards from software design and implementation errors are not controlled since the identical PASS executes on all four machines simultaneously; a single software error (i.e. a common mode failure) could "crash" all four GPC's and, in Garman's words, convert "... the Orbiter to an inert mass of tiles, wires, and airframe..." From this

realization came the idea to include a fifth GPC, which would execute a functionally equivalent but separately designed and implemented flight control system, the BFS. It was (or is) believed that this "software redundancy" would contribute to the overall safety and reliability of the flight software system.¹

In its standby mode, the BFS operates by "listening" to the sensor inputs and some of the outputs from the PASS. Thus the BFS is constantly aware of the current state of the Orbiter and ready to "take over" when necessary (a crew decision). The BFS must stop listening, however, if it detects an inconsistency in the PASS's data processing, so that the integrity of the BFS can not be compromised by a failure in the PASS (a separation of failure modes). On the day of the bug incident, the BFS was tricked into believing that the PASS was compromising its data, and so properly refused to initialize (i.e. start listening to the PASS). The BFS was tricked because the BFS processes (processes can be thought of as individual tasks or programs) and some, but not all, of the PASS processes were out of phase with other PASS processes that were responsible for the "polling" of data from sensors. Thus the BFS regarded all the information from the sensors as garbage. This out-of-phase condition occurred because the BFS processes and some, but not all, of the PASS processes were scheduled based on a cycle counter, while the remaining PASS processes were scheduled based on a separate calculated startup time.² The problem was that this calculated startup time appeared, to the computer, to be in the past because it was compared to a value in a supposedly empty timer queue that unfortunately was not empty. The computer let the "past" start time "slip" into the future (as would an alarm clock set to go off an hour ago) and hence the out-of-phase condition. The queue was not really empty

¹ This is debatable, even among the system developers. The BFS adds significant complexity to the system (as will be discussed), and in fact the present failure was an indirect result of the BFS implementation. Further, this "software redundancy" concept is essentially like the "N-version" approach to achieving software reliability, which has its detractors. Both rely on the assumption that software design and implementation errors for functionally equivalent but independently developed systems are in fact independent; this assumption may not be justified, as discussed in a later section. In any case, the debate should not center on whether or not the development of the BFS system improved reliability and safety. The debate should instead center on whether the costs of BFS development would have been better spent elsewhere, e.g. on further design and testing of the PASS. One wonders to what extent development of the BFS was motivated by uncertainty about the trustworthiness of the PASS, or by an ability to "point the finger" in the case of a disaster involving flight software. As Garman says, "(because of the presence of the BFS) almost everyone involved in the PASS-side 'feels' a lot more comfortable!"

² If you find this incredibly complex and confusing, don't worry - it is. The point here is not to try to understand all the detail, but rather to appreciate the subtlety and complexity of problems that can and do occur in real-time control system software. To me, the PASS/BFS design sounds like an SRM&QA nightmare.

because an isolated data initialization process put a "delay" value in the timer queue that specified when to start the initialization routine. This delay was initially small and did not produce any unwanted side effects. But late in the system development, about one year prior to launch, this isolated value, a single "constant in the code," was increased to a value large enough so that the out-of-phase condition was possible (the value was increased because of a totally unrelated problem, and by this time much of the system testing and integration had been completed). It is significant that much of the problem stemmed from a basic incompatibility between the BFS and the PASS; the PASS is an asynchronous system (i.e. a priority interrupt system) while the BFS is a synchronous system. The PASS system had to be made to "look synchronous" to the BFS, but the emulated synchronism was implemented unevenly (only in "critical" processes), and so the possibility arose for processes to become out-of-phase.

The bug that delayed the initial orbital launch of the Space Shuttle is one particularly well documented example of the types of problems that can occur in real-time software.³ As shown, problems can be caused by seemingly inconsequential changes to seemingly unrelated code segments. Note also that the timing of events is important, yet the timing constraints that must be satisfied for safe operation may be unknown or unclear. The following statement by Garman is appropriate:

"It is complexity of design and process that got us (and Murphy's Law!). Complexity in the sense that we, the 'software industry' are still naive and forge into large systems such as this (the Shuttle) with too little computer, budget, schedule, and definition of the software role. We do it because these systems won't work, can't work, without computers and software."

Garman also notes that the ability to quickly and easily modify software (unlike hardware) is a two-edged sword; this flexibility can reduce costs and lead to better designs, but uncontrolled flexibility can create disasters. Recently, a Shuttle engine test was scrubbed by a computer software check because of an apparently sluggish valve. The valve controls the flow of liquid hydrogen and must be no more than 20% open to prevent a fuel system rupture. According to an Associated Press report in The Huntsville Times (August 5, 1988), NASA had the options of replacing the valve or changing software

³ Many other illustrative examples have been recorded over the past decade in the issues of Software Engineering Notes [Neumann, 1985].

commands to accomodate the way the valve worked (?) during the test. NASA officials decided to change the valve.

3.2. Future Trends

Software controls have many advantages over hardware controls, including greater precision, flexibility, and (perhaps) reliability, as well as allowing a high degree of automation. So software will continue to be an important part of future aerospace systems, and more of the same types of problems can be expected. The Space Station, for example, will depend on the Space Station Information System (SSIS), Data Management System (DMS), Technical and Management Information System (TMIS), and other communications, tracking, and ground support software to perform its important information gathering functions. Further, these software elements will need to interface successfully with each other and with crew members, laboratory experiments, and (life critical) software controlled subsystems such as the Environmental Control and Life Support System (ECLSS).

Perhaps more significant for the future are the new set of software safety problems posed by the incorporation of "advanced software technology" into the Space Station operations software. The use of artificial intelligence and expert system technology is mandated by the Space Station Program office. Consider the following Space Station requirements [NASA, 1988a]:

"(capability for) Growth of artificial intelligence and robotic technology"

"Incorporation of machine intelligence in the form of expert systems, initially for well-defined and structured applications and later for more advanced applications"

"Complex information interfaces of telerobots and autonomous robots. The following are example types of potential information required by the intelligence of robotic devices:

- The location at all times
- The location of obstacles and how to avoid them at all times
- The proper interaction sequence with objects to be manipulated
- Status of itself and its task object
- Its own limitations with respect to its current environment."

These advanced technology requirements either require new software concepts (e.g. artificial intelligence and expert systems) or require new software applications (e.g. control of telerobots and autonomous robots in hazardous space environments). These requirements add significant complexity to the envisioned Space Station software system and add new types of safety and reliability problems that are not found in other real-time control and monitoring systems. Unfortunately, common software engineering management and development practices may be ineffective against these new problems, as they have been against the more traditional problems of the past.

3.3. Failure Costs

Some would argue that serious measures to improve software safety in the manned space program are unjustified because of added costs and limited funds. The space program is, after all, a large-scale research program, so why doesn't NASA just "get on with it," and put the hardware into space without worrying so much about the safety of this or the reliability of that. This seems to be the view of Senator Jake Garn, R-Utah, who flew on Discovery in April 1985. In an August 5, 1988 Associated Press article in The Huntsville Times, Garn was quoted as saying "I think that at this point we probably are being a little bit too cautious because of all the attention (from the news media)." Garn may be right. After all, the potential catastrophic impact from a manned space disaster surely cannot compare with, for example, the accidental firing of a nuclear missile or the core meltdown of a nuclear reactor. This simplistic analysis is flawed, however, because it does not adequately take into account the costs of failure due to lack of appropriate safety measures.

Any rational decision on how much to spend on software safety, and systems safety in general, must consider not only the costs of implementing the safety measures but also the benefits of implementing the safety measures, i.e. the costs of disasters avoided. For the manned space program these benefits can be very great. The public no longer views NASA as simply a large-scale research and development agency. According to John E. Pike, a space program analyst at the American Federation of Scientists, "A lot of the public support and interest (in NASA programs) grows out of the perception that this represents the best of America and our highest aspirations" (The Huntsville Times, August 7, 1988). As a consequence, NASA's public image can suffer greatly because of failures. In a recent survey of 1,223 adults, nearly half lost confidence in NASA after the Challenger disaster, and 60 percent of those still lacked confidence (The Huntsville Times, July 25, 1988). The effect of future

failures on agency funding should be obvious. According to Pike, "If they (NASA) don't get another dozen flights under their belt before another accident, I'd be real surprised if the shuttle ever flew again. If (an accident strikes) this one (the next Shuttle launch), it would pretty well stop the space program." Such high costs of failure suggest that safety measures, including software safety measures, are probably a bargain if they can measurably reduce the chances of failure.

4. SOFTWARE ENGINEERING

4.1. Current State

Software engineering is the art or science or whatever of designing, constructing, and testing computer programs. David Parnas, a noted computer scientist and consultant, wrote a series of eight essays on why the Strategic Defense Initiative would not be trustworthy because of software development difficulties [Parnas, 1985]. These essays were submitted in 1985 along with Parnas' resignation from the Panel on Computing in Support of Battle Management, convened by the Strategic Defense Initiative Organization. In one of these essays Parnas suggests that the state-of-the-art in software engineering is significantly behind that of other more mature engineering disciplines. His claim is supported by three interesting contrasts between software and other engineered products:

- 1) When most engineered products are designed, tested, and sold, it is assumed that the product is "correct" as per functional requirements) and reliable. Finished software, on the other hand, often has significant "bugs" and may be unreliable.⁴ It is usually expected that the software will improve with subsequent versions, but this is not always the case.
- 2) Most engineered products come with an express or implied warranty, but finished software products often come with a specific disclaimer of warranty.
- 3) Designers in more traditional engineering disciplines have been educated to understand and use a variety of mathematical tools, while designers of software are generally uneducated in even the modest tools that are available to software engineers.

Some reasons for the current state of software engineering are given below. The first reason is suggested by Parnas, the others are my personal observations.

⁴ Respected computer scientist Edsger W. Dijkstra [Neumann, 1985], claims that "most of NASA's software is full of bugs." After the successful moon landing in 1969, Dijkstra asked Joel Aron of IBM, who had been responsible for much of the flight software development, how he "got that software to work okay." "Okay?" Aron replied, "It was full of bugs. In one of the trajectory computations, we had the moon's gravity repulsive rather than attractive, and this was discovered by accident five days before count zero."

Software is complex.

One reason for the current state of software engineering is the inherent complexity of large-scale software projects. Large software systems are discrete systems with an enormous number of possible states. Further, software systems usually have few repeatable structures, so that it is not possible to construct a large software system by assembling a large number of small, identical, software modules.⁵ In contrast the design of large-scale integrated circuits, while discrete and having many possible states, is made simpler because many of the structures are repeatable (note that modern computer hardware is vastly more reliable than computer software).

Success in software development may, in fact, be limited by the complexity of the underlying system. In the conclusion to one of his essays, titled The limits of software engineering methods, Parnas discusses the difficulty of developing a trustworthy software system for SDI battle management:

"I am not a modest man. I believe that I have as sound and broad an understanding of the problems of software engineering as anyone that I know. If you gave me the job of building the system (SDI battle management software), and all the resources that I wanted, I could not do it. I don't expect the next 20 years of research to change that fact."

Software is abstract.

Software only exists in the computer, which is very different from the world that people live in. Software can not be seen, touched, or heard, it does not have material strength, and it does not bend, twist, chip, split, or otherwise wear out. This abstract property makes software design appear inefficient to the near-sighted, because design and coding (i.e. "construction") cost about the same but coding produces something tangible (the code can be run on a computer and large piles of results can be printed). Thus there is a strong tendency in software engineering to skip over the design phase to the coding phase. In many other engineering disciplines, however, there is a clear cost advantage to establishing rational and efficient design

⁵ Neumann [1984] quotes from a book called Software and Its Development by Joseph Fox, where Fox discusses the problems of testing Air Traffic Control software systems: "The number of possible paths through these large programs, and the number of possible combinations of states of inputs, data, calculations, and interactions is so large that even in 100 years of use, we will only be beginning to execute the first few percent of the possible paths (through the code). Even after years of real use, there will still be bugs in the program."

methods because an iterative "construct and test" approach is too expensive.

Software design requirements may be ambiguous or incorrect.

This may be just a result of the complexity of software as discussed by Parnas, but it is important enough to deserve separate mention. Software requirements are often written without specific knowledge of the hardware and without complete knowledge of the functional requirements. In contrast a civil engineer, for example, can make logical assumptions about the loads a structure must withstand and the "worst case" combination of loads.

Software engineering is multidisciplinary.

It seems intuitive that a software designer should understand the application area as well as the software engineering discipline (remember that computer science is fundamentally an applied discipline). Unfortunately, these larger-than-life persons are difficult to locate, and in any case would probably demand far too much money. As a result, software is usually written by persons who only know software or who only know the application.⁶

4.2. Modern Approaches

The current state of software engineering should concern those responsible for the safety, reliability, and quality of software. In recent years several software engineering design approaches and tools have been developed to aid the software engineering manager and the software designer. These approaches and tools include: the concept of the software acquisition life cycle; independent verification and validation of software projects; and computer-aided software engineering tools. Although these concepts and methods are not a panacea-they do not guarantee high quality software-they can, if implemented uniformly, be the foundation for other design approaches and techniques aimed at, for example, improving software safety. Uniform implementation of these ideas also allows accurate measurement of the effects of future changes in software design approaches. Each of these design approaches and tools is discussed in the following paragraphs.

⁶ I read a recent letter to the editor that argued for a return to the "good old days" when computer scientists were fundamentally rooted in an area of application (e.g. some type of engineering) and applied computer science principles to their field of interest. This person suggested that the four year undergraduate degree in computer science be eliminated in favor of a two year graduate degree.

The software acquisition life cycle (SALC).

The SALC comprises a heirarchical set of software design, coding, and maintenance tasks, and associated documentation, that define the software acquisition process from initial concept phases through final delivery and maintenance phases. The SALC is a highly structured approach to softwear engineering. Specific requirements associated with each phase are satisfied and baselined before the next phase is begun. Control is exercised at the transitions between phases to assure consistency of requirements and specifications.

The NASA SALC is shown in Figure 1, which is taken from the NASA guidebook titled: Software Verification and Validation for Project Managers [NASA, 1987a].⁷ Loesh [1988] gives a general description of the objectives of the various life cycle phases, which for convenience is repeated below.

- 1) Concept and Project Definition Phase. Ensure software plans, policies and management are appropriately part of early plans and design, i.e. feasible/scoped/costed.
- 2) Software Initiation Phase. Specify a formal system/software definition and consummate the acquisition/development plans (lowest risk delivery agreements).
- 3) Software Requirements Definition Phase. Scope each program, establish engineering procedures and confirm computer/software compatibility (last chance to say what you want).
- 4) Software Preliminary Design Phase. Evaluate the following: 1) are there any major design flaws, and 2) is it consistent with the hardware it is planned to execute to?
- 5) Software Detailed Design Phase. Determine that the module does the right things and will be coded following good engineering principles.
- 6) Software Implementation Phase. Build and show as agreed that the program works correctly.
- 7) Software and System Integration and Testing Phase. Does the system do what was requested?

⁷ This guidebook has been updated recently, and the SALC in the new version may be somewhat different.

ORIGINAL PAGE IS
OF POOR QUALITY

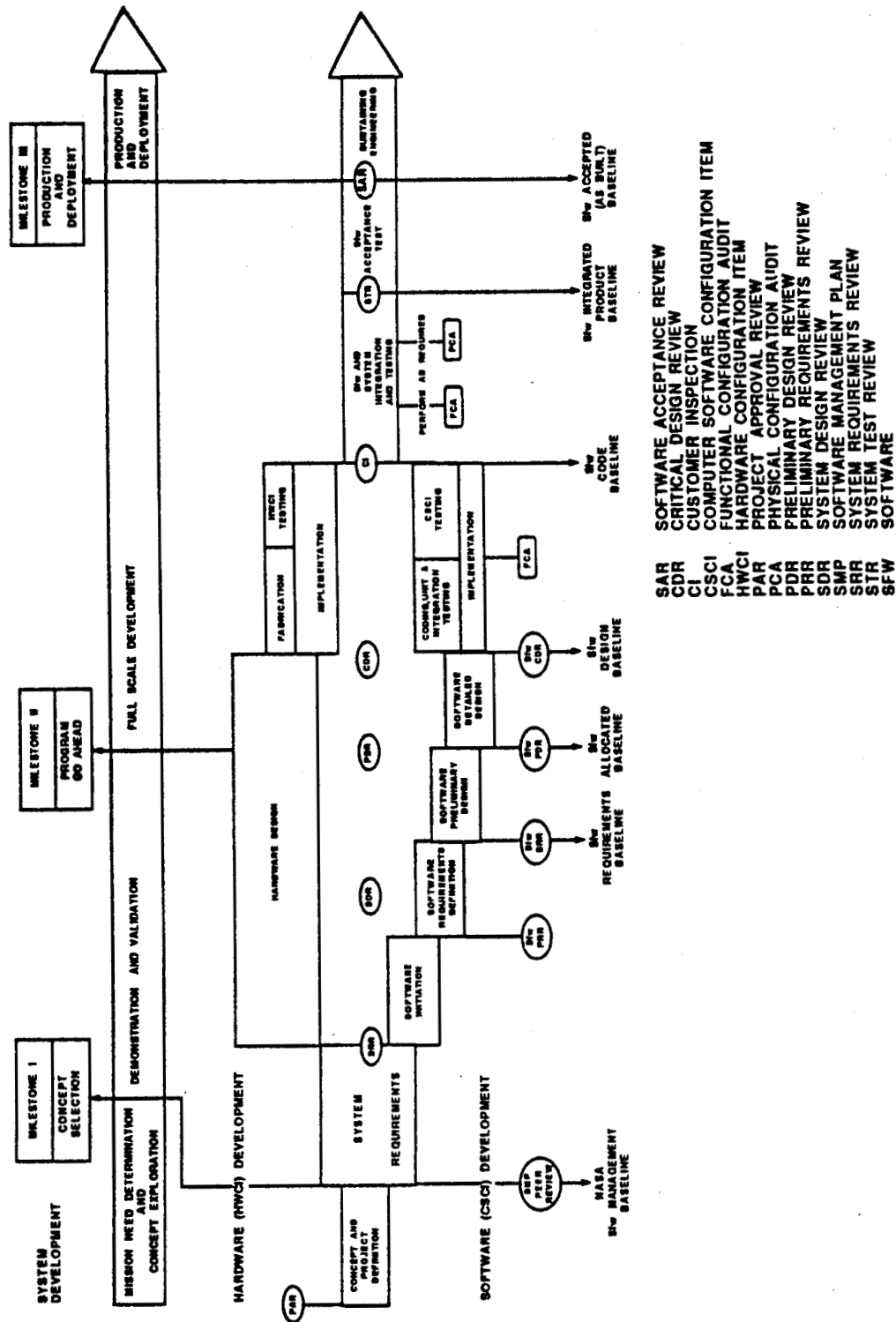


Figure 1 - Software Acquisition Life Cycle

8) Software Acceptance Testing Phase. Get the software and pay the bill.

9) Sustaining Engineering Phase. Fix latent software bugs and produce product upgrades, while ensuring that the software still works correctly. Manage software distribution and archival (including documentation).

It should be obvious that successful implementation of the SALC concept does not come easily and without costs. It has been demonstrated in practice, however, that the life cycle concept is cost efficient when considering costs over the entire useful life of the software (recall the statistics on military software spending). The life cycle model is also thought to improve the general quality of the finished software product (which of course can be linked to cost in terms of sales revenues, disasters avoided, and reduced maintenance). Perhaps the real reason why the SALC works is that actual software coding (implementation phase) is postponed till late in the project. The software managers and designers must think hard in the beginning about requirements, specifications, and associated design issues. This process (and the sheer volume of associated paperwork) allows the designers and implementers enough time to mull over the various design possibilities and choose a good one before committing to code, all the while "appearing" to be productive. Some successful software engineering firms have implemented rigid policies so that on a one-year project, for instance, coding cannot commence until the ninth month! Another important function of the SALC is that it provides the development framework necessary for implementing design tools and procedures aimed at accomplishing more specific software design objectives, e.g. increasing software safety.

Independent verification and validation (IV&V).

Software independent verification and validation is a technical discipline that is in effect during all phases of the software acquisition life cycle. The various IV&V activities and their relations to the SALC phases are shown in Figure 2, taken from the previously cited NASA guidebook. As shown, IV&V activities include designing and executing tests based on information from requirements and specification documents, design documents, and code audits. These tests are meant to ensure that the final delivered product meets the stated requirements and is of high quality. In addition to formal testing, IV&V activities also assure that each software end item, whether a individual module, a partially integrated subsystem, or a fully integrated system, satisfies its corresponding requirements and specifications as prescribed during the hierarchical design process (a definition of



Figure 2 - Software IV&V Activities

"validation"), and that the products of each phase of the SALC satisfy the requirements and specifications of the preceeding phase (a definition of "verification"). It seems logical that IV&V staff should be mostly seasoned, experienced personnel (who knows all the "tricks of the trade"); this is no place for the green programmer with little practical experience.

The "independent" in IV&V refers to the relationship between the project management (NASA), the development contractor, and the IV&V contractor. The quality of the IV&V effort and hence the quality of the delivered software is thought to increase with the degree of independence between these players. Minimal independence is achieved by a development contractor who uses the development staff for IV&V (really just V&V), while maximal independence is achieved by a wholly separate IV&V contractor that communicates with the developer and with NASA by separate channels. A truly independent verification and validation contractor provides an unbiased critical review and "second opinion" of the software. Independence assures that the testing and design reviews will not be polluted by the day-to-day design activities, and that IV&V personnel will not be competing for resources with development personnel. A higher degree of independence does, of course, add significantly to the up-front project cost. These costs should properly be weighed against the criticality of the software when deciding on the degree of independence. This trade-off analysis should also consider the non-trivial costs of the sustaining engineering phase (maintenance), which are presumed to decrease with an increase in independence.

Computer aided software engineering (CASE).

CASE tools are programs to help write programs. CASE tools emphasize a systems approach to software engineering, as opposed to a ad hoc "software crafting" approach. Thus CASE supports the life cycle model of software development, and can make the transition to SALC easier by assisting the programmer in various tasks associated with the life cycle phases. The following are examples of CASE tools that either exist or are under development: program design languages and specification languages; test design aids; symbolic debuggers; automatic code generators; and configuration management aids. These tools are usually graphical and highly interactive. Other CASE tools implement the structured approach to requirements specification that is growing in popularity [DeMarco, 1979].

The future Space Station Software Support Environment (SSE) is an ambitious effort by NASA to supply Space Station software developers with a consistent and comprehensive set

of CASE tools. Consider just a few of the stated requirements for the SSE, taken from the SSE system functional requirements specification [NASA, 1988b]:

"The SSE system shall support project initiation and control, project management, metrics collection and analysis, project planning, budgeting, resource allocation and accounting, scheduling, lessons learned gathering and analysis, and performance management."

"The SSE system shall provide the capability to determine the complexity and criticality of a proposed change and to prioritize proposed changes based upon those metrics."

"The SSE system shall support computer-aided preliminary design for SSP operational software and SSP prototypes, models and simulations."

"The SSE system shall support software checkout and verification from the unit level, to the subsystem level, to the complete software system, and finally to integrated software and hardware systems."

And the list of requirements goes on and on. Once complete, the SSE will be a state-of-the-art programming environment for the development of all Space Station software. Further, the SSE will support the development of new software engineering tools and methods as well as their rapid integration into the development life cycle. Some of these new tools and methods could address timely issues in software safety.

5. SOFTWARE SAFETY

Software by itself is harmless. When software is embedded in a hardware system, however, say for a real-time control application, inappropriate software actions or inactions can lead to hazardous system states with catastrophic potential. Nevertheless, modern software development efforts do not normally consider these software safety concerns. Safety is not usually included in the software requirements specification, and IV&V activities address software "quality" issues but not necessarily software safety.

Comprehensive and generally applicable software safety methods, tools, and techniques do not currently exist. Software safety can be achieved only through software design (as opposed to redundancy as in hardware). Unfortunately, there are few adequate mathematical tools for analyzing software designs, and few adequate metrics for measuring software safety. It is unlikely that a "safety factor" concept can be applied to software, and at the present there is no comprehensive "building code" that will assure software safety.

Although technology lags behind, there is a strong institutional motivation for developing and implementing software safety methods. Applicable government standards and new project requirements are now addressing software safety issues. This new focus is especially noticeable in the Military safety program requirements and the program requirements for the Space Station.

The Military standard on system safety program requirements [DoD, 1987] includes a new task section (as of July 1987) on software hazard analysis requirements. This set of requirements emphasizes a systems approach (which is necessary, of course) that includes analyses of the safety critical hardware/software interfaces. Thus information from hardware safety analyses (hazard analyses, FMEA's) are used as input to the software safety analysis, and leads to the concept of a Safety Critical Computer Software Component (SCCSC). These SCCSC's or sets of interfacing SCCSC's can be targeted for detailed investigations to determine, for example, whether certain hazardous events can occur as a result of software actions or inactions, or if adequate software safety controls are present. This standard appropriately supports the life cycle software development model, and includes software hazard analysis requirements for each of the major SALC phases.

A variety of Space Station Program (SSP) documents specify software safety requirements. The software product assurance requirements [NASA, 1988a] state that the provider "shall plan, document, and implement a software safety process." The MSFC level C Space Station software management plan [NASA, 1987b] requires a software product assurance plan that emphasizes "software fault tolerance and software failure mode and effects analysis for all critical software packages." The instructions for SSP hazard analysis preparation [NASA, 1987c] includes software hazard analyses in the scope of work, and describes a software hazard analysis approach that includes the use of software fault tree analysis.

The cited Military standards and NASA documents say what to do but not how to do it (which is appropriate for documents of this type). As indicated above, the current technology for the "how" part is probably inadequate for implementing the "what" part. Recent concern about software safety has, however, spurred the investigation of new software safety tools, techniques, and methods [Leveson, 1986]; these approaches are summarized below. Most investigations have borrowed existing hardware safety analysis techniques and applied them to software, and most have applied the techniques to relatively simple test systems. Although in an early stage of development, some of the techniques show promise, and the continued development of these and other software safety approaches should be supported in order to meet the software safety needs of the future.

Software fault tree analysis (SFTA).

The application of fault tree analysis to software was investigated by Leveson and Harvey [1983], Leveson and Stolzy [1983], Leveson [1984], and Cha, Leveson, and Shimeall [1987]. SFTA uses a knowledge of the programming language to build a fault tree that describes all possible software paths leading to a particular hazardous event. Successful paths (those which produce the hazardous event) must be eliminated or controlled or justified as very unlikely to occur in practice. In Leveson and Harvey [1983], SFTA was applied successfully to a flight and telemetry control program for a University of California, Berkely, spacecraft (approximately 1250 lines of assembly language code). The SFTA identified a logic error whereby two sun-pulse interrupts within 64 ms of each other could crash the microprocessor and render the spacecraft useless (such a condition is highly unlikely, but one of the "sun-pulses" could be artificially produced by gamma radiation). A simple software check could eliminate this hazardous condition.

SFTA is practical when there exists only a limited number of safety critical failures, a condition that is apparently often satisfied. SFTA is thought to be helpful because it provides a structured approach to thinking about software problems from a safety perspective. It also may be possible to automate the generation of the software fault trees by examination of the code. A disadvantage of SFTA is the difficulty of including timing related failures in the analysis.

Petri nets.

The application of Petri nets to software safety analysis was investigated by Leveson and Stolzy [1987]. A Petri net is a dynamic system modeling approach consisting of places, transitions between places, and a set of tokens marking the places and defining the current system state. Software systems can be modeled by Petri nets and the Petri nets can be executed, which consists of "firing" the tokens along the enabled transitions toward a new set of places. Repeated firing of the Petri net models the dynamic system behavior. The advantage over static analysis approaches (e.g. SFTA) is that important timing properties of programs can be investigated. Run-time faults and failures can also be incorporated into the analysis, and there is significant opportunity for computer-aided analysis. In the paper by Leveson and Stolzy a Petri net model was applied to a simple railroad crossing controller.

Sneak software analysis.

In sneak software analysis the program is converted into flow diagrams using electrical symbols, and is then analyzed to detect certain logic errors such as undefined variables. Leveson [1986] points out that much of this information is provided by good compilers, and that furthermore sneak software analysis is more a reliability than a safety technique since it attempts to identify all faults. She further notes that it is unlikely that many significant faults will be found this way (of the type that were discussed previously), and draws a parallel between trying to find significant errors using sneak software analysis and "trying to find the errors in a book by checking the grammar."

N-version programming.

N-version programming was investigated by Avizienis [1985] as an approach to fault-tolerant computing. In this approach N different programs are developed from the same functional specifications. The developers should be as independent as possible, and should use different tools and

compilers. These N program versions are executed simultaneously and the results are polled and compared by a voting program, with the correct answer taken as the majority answer. Sophisticated methods have been devised to provide for communications between the various versions and the voting program, and hence to ease development efforts. The approach depends, however, on the assumption that software design and implementation errors associated with the different versions are independent (a high degree of dependence of errors could create situations in which the majority of program answers agree but are all wrong). Unfortunately the approach may be limited by this assumption, as a recent study suggests that the assumption of independence may not be justified [Knight and Leveson, 1986] (consider the special problem of dependence because of common specification errors). Further study is needed to evaluate thoroughly the assumption of independence before N-version programming can be used widely with confidence.

6. CONCLUSIONS AND RECOMMENDATIONS

Large-scale software systems are often used to control real-time flight and ground operations on manned space missions. This trend is expected to increase because of the emphasis on automation and advanced software technology. There is ample evidence that suggests these large software programs will contain logic errors that may under certain situations lead to hazardous conditions. Thus it is important to emphasize safety-related approaches to software development throughout the useful life of the software, beginning at requirements development. Software safety engineering is a young discipline, however, and the necessary analysis and development tools are only beginning to be developed. It is a tough problem, since software is highly complex, abstract, and does not lend itself easily to mathematical analysis.

SRM&QA should take a leadership role in the development and refinement of software engineering methods, techniques, and tools that can help to improve software safety, reliability, and quality. This is especially important for software safety, because while software reliability and quality have many proponents, software safety is largely neglected because it is separated from functional requirements and traditional performance measures. SRM&QA should strive to build and maintain a knowledge base of sound software engineering practices that it can distribute to contractors for use in project development.

Several specific recommendations follow, not necessarily in any order.

SRM&QA should continue to develop its software product assurance function and to promote SPA methods.

The SPA activities (including IV&V) are the backbone of any effort to improve software safety, reliability, or quality. SRM&QA should logically be responsible for the bulk of NASA SPA activities and in fact for the bulk of NASA IV&V contracting and monitoring. NASA should consider a software project management structure that provides for internal independence of development and IV&V engineers and managers. SRM&QA should develop a consistent and rational method for determining the necessary degree of independence of the IV&V contractor based on software criticality, total life cycle cost, and other factors.

SRM&QA should develop a software hazard analysis methodology.

This development effort may include an evaluation of applicable Military standards and other guidebooks and instruction manuals with regard to NASA's needs; an evaluation of personnel requirements including skill and education levels; identification and evaluation of existing techniques and computer-based tools for software hazard analysis; development and evaluation of new techniques and computer-based tools for software hazard analysis; and investigation and development of recommended software design methodologies that encourage the production of safe software (i.e. safety metrics and analysis techniques, and "building codes" for safe software development).

SRM&QA should explore the possibility of developing an integrated computer-based environment to support all SRM&QA management and engineering activities.

This computer-based environment would be graphics-based, would include a variety of tools specific to SRM&QA activities, and would interface directly or indirectly with the Technical and Management Information System (TMIS), the Software Support Environment (SSE), the Space Station Information System (SSIS), and the Data Management System (DMS). The potential benefits of this development effort include improved efficiency and quality of work, and expedited flow of information between SRM&QA, other NASA organizations, and contractors.

SRM&QA should develop and investigate methods, techniques, and tools for evaluating the safety, reliability, and quality of software systems that use advanced software engineering technology.

Specifically, SRM&QA should investigate special safety, reliability, and quality problems (and their possible solutions) created by the use of artificial intelligence and knowledge based expert system technology in critical software systems. Use of this technology (or provision for its use) is a requirement of the Space Station Program, and it is not clear, for example, whether current software engineering practices apply or whether anyone even knows how to debug an expert system or artificial intelligence program.

7. REFERENCES AND SELECTED BIBLIOGRAPHY

1. Avizienis, A., "The N-Version Approach to Fault-Tolerant Software," IEEE Trans. on Software Eng., Vol. SE-11, No. 12, Dec. 1985.
2. Cha, S., Leveson, N., and Shimeall, T., Safety Verification of Ada Programs in MURPHY, University of California at Irvine, UCI Technical Report 87-27, 1987.
3. DeMarco, T., Structured Analysis and System Specification, Prentice Hall, Englewood Cliffs, N.J., 1979.
4. DoD, Software System Safety Handbook, Air Force Handbook AFISC SSH 1-1, Sep. 1985.
5. DoD, Military Standard, System Safety Program Requirements, MIL-STD-882B, July 1987.
6. Garman, J., "The 'Bug' Heard 'Round the World," Software Engineering Notes, Vol. 6, No. 5, Oct. 1981.
7. IEEE, The Small Computer Revolution, Proceedings COMPCON, IEEE Computer Society Press, Fall 1984.
8. Jahanian, F., and Mok, A., "Safety Analysis of Timing Properties in Real-Time Systems," IEEE Trans. on Software Eng., Vol. SE-12, No. 9, Sep 1986.
9. Knight, J., and Leveson, N., "An Experimental Evaluation of the Assumption of Independence in Multiversion Programming," IEEE Trans. on Software Eng., Vol. SE-12, No. 1, Jan. 1986.
10. Leveson, N., "Software Safety in Computer Controlled Systems," IEEE Computer, Feb. 1984.
11. Leveson, N., "Software Safety: Why, What, and How," Computing Surveys, Vol. 18, No. 2, June 1986.
12. Leveson, N., and Harvey, P., "Analyzing Software Safety," IEEE Trans. on Software Eng., Vol. SE-9, No. 5, Sep. 1983.
13. Leveson, N., and Stolzy, J., "Safety Analysis of Ada Programs Using Fault Trees," IEEE Trans. on Reliability, Vol. R-32, No. 5, Dec. 1983.

14. Leveson, N., and Stolzy, J., "Safety Analysis Using Petri Nets," IEEE Trans. on Software Eng., Vol. SE-13, No. 3, March 1987.
15. Loesh, R., Software Verification and Validation, Training Course Notes, System Technology Institute, Inc., Rev. C6, 1988.
16. NASA, Software Verification and Validation for Project Managers, Version 0.1, Safety, Reliability, Maintainability, and Quality Assurance Publication D-GL-13, March 1987a.
17. NASA, Level C Space Station Software Management Plan, George C. Marshall Space Flight Center, SS-PLAN-0006, 1987b.
18. NASA, Instructions for Preparation of Hazard Analyses for the Space Station, JSC 30309, Space Station Program Office, 1987c.
19. NASA, Space Station Program Definition and Requirements, JSC 30000, Lyndon B. Johnson Space Center, Houston, Texas, 1988a.
20. NASA, Space Station software Support Environment System Functional Requirements Specification, LMSC F255416, 1988b.
21. Neumann, P., "Letter from the Editor," Software Engineering Notes, Vol. 10, No. 5, Oct. 1985.
22. Neumann, P., "On Hierarchical Design of Computer Systems for Critical Applications," IEEE Transactions on Software Eng., Vol. SE-12, No. 9, Sep. 1986.
23. Parnas, D., "Software Aspects of Strategic Defense Systems," Communications of the ACM, Vol. 28, No. 12, December, 1985.

N89-21755

1988

NASA/ASEE SUMMER FACILITY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

THE KEYHOLE REGION IN VPPA WELDS

Prepared by:	Daniel W. Walsh Ph.D.
Academic Rank:	Associate Professor
University and Department:	California Polytechnic State University, San Luis Obispo Metallurgy and Materials Engineering
NASA/MSFC:	
Laboratory:	Materials & Processes
Division:	Process Engineering
Branch:	Metals Processes
MSFC Colleague:	Arthur C. Nunes Jr., Ph.D.
Date:	August 8, 1988
Contract No.:	NGT 01-002-099 The University of Alabama

ABSTRACT

Title: Physical Processes Associated With the Weld Keyhole; Weld
Macrosegregation and Arc Character

Daniel W. Walsh

Cal Poly San Luis Obispo

The morphology and properties of the Variable Polarity Plasma Arc (VPPA) weld composite zone are intimately related to the physical processes associated with the keyhole. This study examined the effects of microsegregation and transient weld stress on macrosegregation in the weld pool. In addition the electrical character of straight and reverse polarity portions of the arc cycle were characterized. The results of the former study indicate that alloy 2219 is weldable because large liquid volumes are available during latter stages of weld solidification. Strains in the pool region, acting in conjunction with weld microsegregation can produce macrosegregation great enough to produce radiographic contrast effects in welds. Mechanisms of surface copper enrichment were identified. The latter study has demonstrated that increased heat is delivered to workpieces if the reverse polarity proportion of the weld cycle is increased. Current in the straight polarity portion of the welding cycle increases as the reverse cycle proportion increases. Voltage during reverse polarity segments is large.

Acknowledgement

Thanks to all who made this summer experience worthwhile and enjoyable. Special thanks to Ernestine Cothran and Mike Freeman for a well run program, thanks to Billie Swinford and Missy for taking their turn to "watch" me this summer. The list of NASA and contractor employees who deserve my thanks could form the nucleus of a third political party. Risking a sin of omission, these people are Benji Swain, whose constant interest, and insight helped make this summer successful, Dave Newman, Jeff Ding, Carolyn Kurgan, Jeff Norris, Frank Zimmerman, Bill McGee, Wayne Owen, Chip Jones, Rick Underwood, the entire staff of X-ray (please forgive me, I know how many loops you all willingly jumped through for me), Terry Craig, Mr. Craig, Tim Tittsworth (missed your cow) Angie, Ann Whitaker, Ph.D. Paul Gill, Svelte Marty Martinez, and Ronnie (three point land) Renfro. Despite my efforts neither Ken Swaim (my electronics guru) or Sam Clark (diety of welding) have recognized that the PAC 10 is the power conference in the country. I would also like to thank Joe Bucher and Richard Venable for their help, and Tim Vaughn and Dan Drinnan for their helpful comments. A special thanks to Joe Montano and Wendell Deweese, their help was invaluable and much appreciated. Thanks to James Coston and Gretchen Perry for their efforts. Thanks to Ernie Bayless for his interest and comments. Thanks to the a.m. discussion group (Joe Gregory and Art Nunes) for caustic comment cathartic coffee. I would like to thank Bertha Gildon for her help through the summer and for typing this manuscript. Thanks to NASA ASEE for fostering this program. A special thanks to Art Nunes, the personification of the NASA colleague concept. His constant interest, ready insight, and friendship made this summer, as last, a rewarding experience. Finally thanks to my wife and family whose patience and perseverance made this summer work possible and whose expenses make it mandatory.

INTRODUCTION AND OBJECTIVES

The physics associated with the VPPA process is interesting for three reasons. First, though more information is collected each day, the keyhole and pool are not well understood. Indeed the VPPA keyhole process has not been as well characterized as other keyhole processes. Second, fluid flow and heat transfer fix the distribution of microstructure and properties in the weld composite region. Finally fluid flow in the pool and static forces on the pool determine bead morphology. Fluid flow in the weld pool was the subject of intense study during the summer of 1987⁽¹⁾. During this summer program, pool solidification, the electrical character of the arc, and the morphology of VPPA weld beads were investigated. The steady state keyhole is shown in Figure (1). Energy is transferred to the work via a heated plasma directed by the plasma torch. Much of the energy transfer is accomplished by convective mechanisms in the hot, directed efflux plasma and by radiation from an arc that is buried in the keyhole region. In VPPA welding, the arc polarity switches many times a second, but the work is principally anodic. The oxide film associated with Al 2219 is disrupted by the sputtering action of argon ions during the reverse polarity portion of the cycle, and the arc is stabilized. A majority of the heat transfer occurs at the leading edge of the keyhole, where the anode spot is located and the plasma is directed. In this location the liquid is thin, and the thickness uniform regardless of depth in the keyhole. Around the keyhole, until the longitudinal centerline is passed the fluid remains thin. The thickness is not as uniform, becoming slightly larger deeper in the keyhole. In the rear, a croissant shaped fluid pool exists.

Recent success in joining HP-9-4-30 spurred interest in identifying the portions at energy delivered to welds during the straight or reverse periods of the VPPA welding cycle. The character of the welding arc differs drastically during the straight and reverse polarity periods in the cycle. As a rule, voltage increases but current falls during reverse polarity. Two key items examined for effect on power production were cathode emission efficiency, and altered arc resistance.

Recently much effort has been directed towards discerning the nature of a linear x-ray indication found in production welds on the external tank. The indication parallels the welding direction, and mimics dangerous lack of fusion or cracks so well that it triggers a repair response. However, when repaired, no defect is found. Nunes and Ding⁽²⁾ have identified the cause of the indication, but here-to-fore no mechanism of formation has been established. This program established that mechanism. In order to understand the arguments presented a knowledge of weld solidification is required.

Pertinent Solidification Theory

Weld solidification theory can be used to explain the amounts and distribution of constituents in Al-2219 VPPA weld microstructures.

Figure (2) depicts the Al-Cu system, in our range of interest. The nominal composition (C_o) is marked as 6.3% Cu, representing nominal 2219. Under conditions of equilibrium, a liquid of C_o begins to form solid at the intercept of C_o with the liquidus line. The composition of the first solid to form is defined by the temperature at that intersection. With continued cooling the composition of both the solid and the remaining liquid adjust, by diffusion, so that at any temperature the entire solid volume has a composition defined by the intercept of the isotherm and the solidus, and the entire volume of the remaining liquid has a composition defined by the intercept of the isotherm and the liquidus. Clearly, the assumption of uniform composition in the phases implies that the diffusion distances are much less than

Egn. 1

$$\sqrt{D t_f}$$

where D is the limiting diffusion in the system, and t_f is the solidification time. At any temperature, invoking mass conservation.

Egn. 2 $F_s C_s + f_l C_l = C_o$

thus

F_s = fraction solid
 C_s = solid conc. in solid
 f_l = fraction liquid
 C_l = solute conc. in liquid

Egn. 3

$$d(F_s C_s) + d(f_l C_l) = 0$$

Egn. 4

$$\frac{d C_s}{(1-k_o) C_o} = \frac{d f_s}{1-f_s(1-k_o)}$$

Now $C_s = K_o C_l$, where $K_o = C_s/C_l$, the equilibrium distribution coefficient K_o is the ratio of the C_s and its equilibrium C_l . In many systems K_o is a function of temperature because the solidus and liquidus lines are curved. In the Al-Cu system these lines are nearly straight and K_o is constant over a wide range of temperatures. Now $dC_s = K_o dC_l$, and for all fractions solidified, $F_s = 1-f_l$.

Eng. 5

$$\frac{C_l}{C_o} = \frac{1}{1 - (1-k_o) F_s}$$

For a 6.3 Cu alloy we can calculate the fraction solidified prior to the production of eutectic. (at the eutectic $C_l = 33.2\%$ Cu).

Egn. 6
$$F_s = \left(\frac{C_o}{C_e} - 1 \right) / (1 - k_o)$$

or $F_s = 0.975$. Thus the weight fraction of eutectic is 0.025. This behavior is shown in solidification path 1 in Figure 2. A schematic depicting the conservation of solute for this "equilibrium solidification" is shown in Figure 2. Note that the dearth of solute in the solid is exactly equal to the excess solute in the liquid at all f_s . Note further that dC_s/dx and dC_l/dx are both zero. Equilibrium of the sort described above does not occur in nature, particularly not in welds. The heat of fusion that must continually be dissipated to the surroundings prevent attainment of equilibrium, but the redistribution of solute required in both the solid and liquid poses an even greater obstacle. Because adjustments to the solid composition can occur only by diffusion, the adjustment is sluggish and limiting. Postulating no diffusion in the solid, but complete mixing in the liquid (dC/dt) $_x = 0$ in solid, (dC/dx) $_x = 0$ in liquid), we can calculate a corrected volume of eutectic. Equilibrium is still maintained, but only at the microscopic level. In regions a few atoms thick on either side of the solid-liquid interface free energy is minimized if C_s and C_l are as predicted by the phase diagram. The behavior of a 6.3% Cu alloy, solidifying under these constraints, is shown as 2 in Figure 2. Notice that the average composition of the solid, at any temperature, is intermediate to the solid in equilibrium with liquid of nominal composition and the solid composition at the interface, C_s^* . Assuming a rectangular volume of liquid with unit cross section, solidifying by the passage of a planar front

Egn. 7
$$(C_s - C_s^*) dF_s = (1 - F_s) dC_l$$

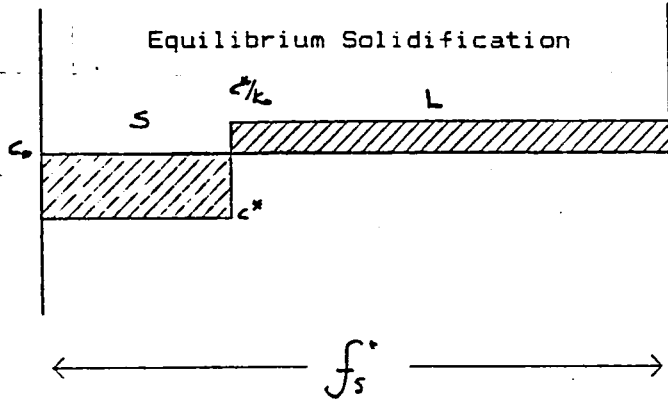
or

Egn. 8
$$(C_l - k_o C_l) dF_s = (1 - F_s) dC_l$$

therefore

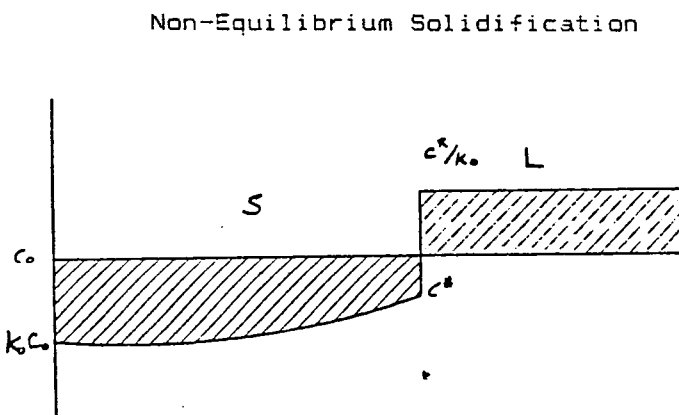
Egn. 9
$$\frac{dF_s}{(1 - F_s)} = \frac{dC_l}{C_l(1 - k_o)}$$

PERTINENT SOLIDIFICATION THEORY



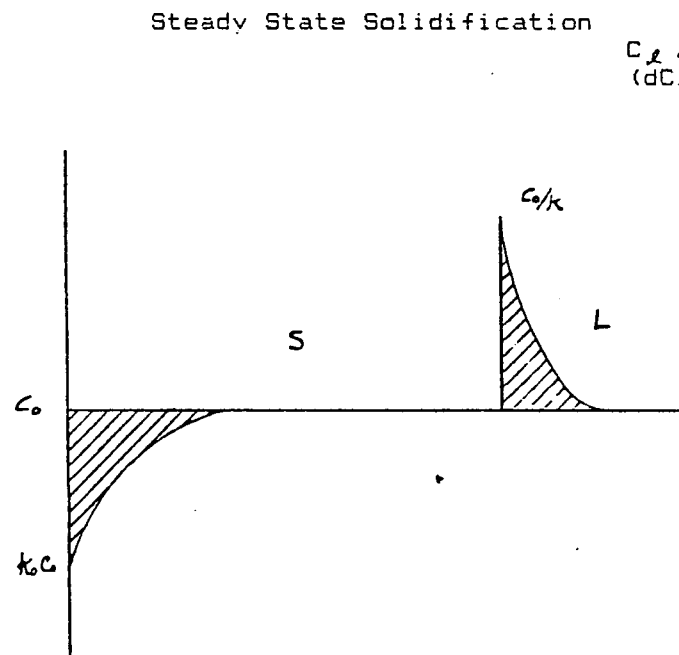
$$C_s / C_o = \{ 1 - (1 - k_o) F_s \}$$

Complete Mixing both S,L



$$C_s / C_o = (1 - F_s) (k_o - 1)$$

$$(dC/dt)_{x,s} = 0 \quad (dC/dx)_L = 0$$



$$C_s / C_o = \{ 1 + [(1 - k_o) / k_o] e^{(-RX/D)} \}$$

$$(dC/dt)_{x,s} = 0 \quad \text{Diffusive Mixing in L}$$

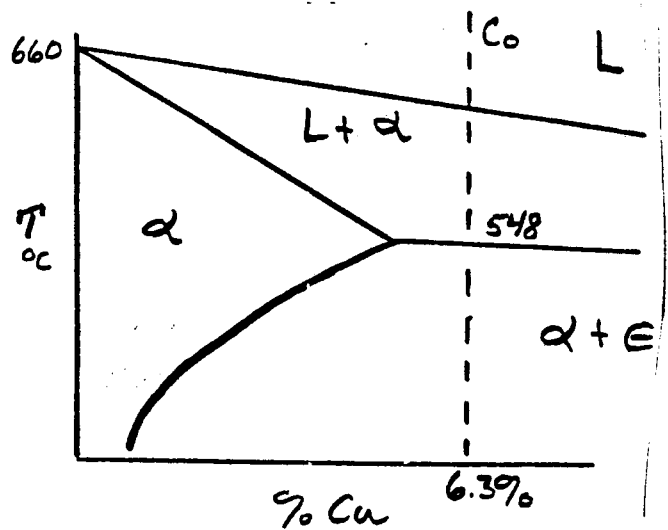


Figure 2.
XXIX-4

ORIGINAL PAGE IS
OF POOR QUALITY

Integrating we find

$$\text{Egn. 10} \quad -\ln(1 - F_s) = \left(\frac{1}{1 - k_0}\right) \ln \frac{C_e}{C_0}$$

thus

$$\text{Egn. 11} \quad C_e = (1 - F_s)^{(k_0 - 1)} C_0$$

and

$$\text{Egn. 12} \quad C_s^* = k_0 C_0 (1 - f_s)^{k_0 - 1}$$

This behavior yields 0.13 volume fraction of eutectic, much more than under equilibrium conditions. The increase is caused by the change in the solute distribution in the solid. Figure 2 shows this effect schematically. Composition is shown as a function of distance (which corresponds to volume fraction). More eutectic forms because the initial solidified cell composition, and thus the average composition when the eutectic starts to form is less than under equilibrium conditions.

Steady state solidification can be produced by assuming diffusive mixing in the liquid. Given a rectangular volume of unit cross section, subject to diffusive flux the change in concentration in any differential volume as a function of time can be calculated. If two planes are separated by dx the flux into the volume, across the first plane is

$$\text{Egn. 13} \quad J_1 = -D \left(\frac{dc}{dx} \right)_1$$

and the flux across the second is

$$\text{Egn. 14} \quad J_2 = -D \left(\frac{dc}{dx} \right)_2$$

and the net flux can be found as the difference between J_1 or J_2 .

$$\text{Egn. 15} \quad \text{Now } \left(\frac{dc}{dx} \right)_2 = \left(\frac{dc}{dx} \right)_1 + \left(\frac{d^2c}{dx^2} \right) dx$$

thus

$$\text{Egn. 16} \quad J_{net} = D \left(\frac{\partial^2 c}{\partial x^2} \right) dx$$

$$\text{Egn. 17} \quad J_{net}/\partial x = dc/dt$$

$$\text{Egn. 18} \quad dc/dt = D \partial^2 c / \partial x^2$$

It is instructive to examine a rectangular volume, solidifying from one end, Figure 2. Solidification in this case is subject to the constraints that $dc/dt = 0$ or that there is diffusive mixing in the liquid. The first solid to form has composition $K_0 C_0$. The remaining liquid is enriched in solute ($K_0 C_0 < C_0$). Because there is no mechanical mixing in the liquid a solute concentration gradient is developed in the liquid adjacent to the solidifying interface. Again, at all times solute is conserved. The loss of solute in the solid corresponds to the solute gain in the liquid. As shown previously the loss of solute from a volume element in the gradient region is

$$\text{Egn. 19} \quad J_D = D \frac{\partial^2 c}{\partial x^2}$$

However, the moving interface sweeps up solute, and the flux by interface migration is

$$\text{Egn. 20} \quad J_m = R \frac{dc}{dx} \quad R = \text{rate of solidification}$$

When C_L at the interface reaches C_0/K_0 a condition of dynamic stability is established and

$$\text{Egn. 21} \quad R \frac{dc}{dx} + D \frac{\partial^2 c}{\partial x^2} = 0$$

This equation has a solution of the form

$$\text{Egn. 22} \quad C_x = C' e^{-(R/D)x} + C_0$$

where $C' = \frac{C_0}{K_0} - C_0 = C_0 \left(\frac{1-K_0}{K_0} \right)$

and $X' = \text{distance ahead of the solid-liquid interface}$

thus

$$\text{Egn. 23} \quad C_x = C_0 \left[1 - \left(\frac{1-K_0}{K_0} \right) e^{-(R/D)x'} \right]$$

Materials and Procedures

The experimental portion of this study was comprised by two major programs:

- 1) Determination of the events that lead to the formation of linear enigmas in Al-2219 VPPA welds.

- 2) Determination of the electrical character of the VPPA arc during the straight polarity portion of the cycle vis-a-vis the reverse segment.

In the enigma program, several series of welds were made. First, a group of modified production welds were made on 0.500 inch thick, 10" x 24" aluminum 2219 plates. Each plate was welded in the vertical position, using a production schedule for plates of this thickness to make bead on plate welds. The first pass was a fully penetrating keyhole 20" long. the second (fill) pass was made, terminating 1" prior to the root pass. Six subsequent autogenous passes were made, the next two terminating after 14" of weld, the final four after 9" of weld. Three plates of this type were made, one with standard production parameters, one with the same parameters, but a 15ms/8ms straight/reverse cycle (vis-a-vis the typical 19ms/4ms), and one with increased torch standoff. Each weld was sectioned to provide samples from each region for optical microcopy, scanning electron microscopy, and radiography. Samples examined in the x-ray were wrapped in lead to minimize anomalous surface effects (backscatter). Plate sections were also characterized using the Computer Aided Tomography system. A fourth plate was welded using the parameters for the first plate. However, instead of overlaying the cover passes, each subsequent pass was moved over one half bead width, (alternating sides) producing a "Nydra" like effect.

Several welds were made while using the UV Nitrogen Laser Vision system to record the appearance of the posterior of the keyhole. These welds were made in 0.375" plate in the vertical position. The weld toe was examined in detail, partial melting and local plate distortion were recorded.

Several stationary VPPA welds were made on 1" thick Al 2219 in the vertical position. Three different arc durations were used, one, tow, and five minutes. These welds were examined with stereography and scanning electron microscopy. Finally, several two pass welds were made on 0.500 inch plate while the plate was clamped in a vise. This clamping force put the plate in compression over central region ten inches long. The behavior of the weld was monitored, and several transverse sections were examined using optical microscopy.

During the course of the second program twelve 0.500 inch aluminum 2219 plates were welded, all in the vertical position. In addition four 0.250" Inconel 718 plates were welded, all in the flat position. Table 1 lists the welding parameters used in this program. In the course of the study welding amperage, welding current, plasma gas pressure, standoff, and straight/reverse polarity distribution were varied. Welding voltage and welding current were measured using photographs of stored oscilloscope traces. The welding operation was abruptly terminated to retain a fossil record of the keyhole. The welds were sectional and characterized by both an optical comparator and after polishing an optical microscope. An energraphics plotting routine was used to generate expressions that described weld contour. The weld region was characterized by microhardness transverse.

RESULTS AND DISCUSSION

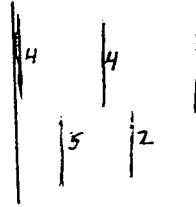
Enigma Formation

Figure 3 shows the weld composite zone. A region of complete melting (fusion zone), a partially melted region, and a region where the base metal character is altered only by solid state reaction is evident. Figure 3A contrasts the microstructure of the base metal far removed from the weld and microstructure of the fusion zone. Using equations developed in the introduction, the volume fraction of eutectic constituent is calculated as 2.5% for the base metal, but 13% for the weld region. These calculated values correspond well with the amount of eutectic seen. Under fully equilibrium conditions, the base metal is composed of the aluminum rich solid solution of Cu in a FCC structure (α) and CuAl₂. However, because the 2219 alloy has a complex thermomechanical history, the α phase is super saturated in Cu. The alloy is cold-worked and aged to foster the precipitation of metastable phases on dislocation networks. As the alloy is overaged, metastable phases transform to CuAl₂, which grows and becomes visible to optical microscopy. Throughout the matrix, before overaging, the 2.5% eutectic present is in a divorced morphology. Globules of CuAl₂ appear as a distinct phase in the structure rather than as a eutectic mixture of two phases. The eutectic freezes in a "divorced" fashion because the α component freezes epitaxially on the primary α surrounding the eutectic liquid, leaving the tetragonal CuAl₂ *laves* to form terminally.

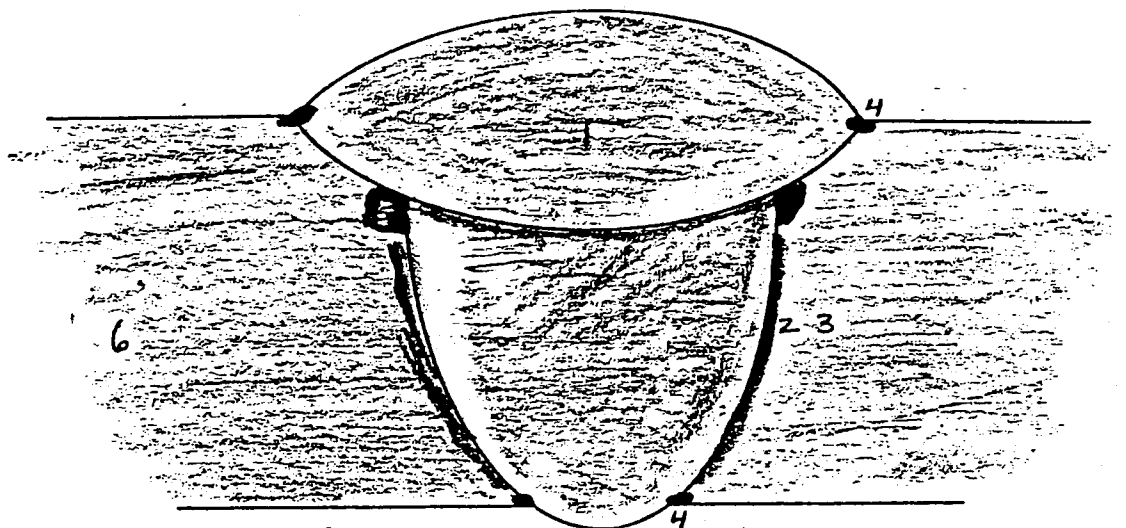
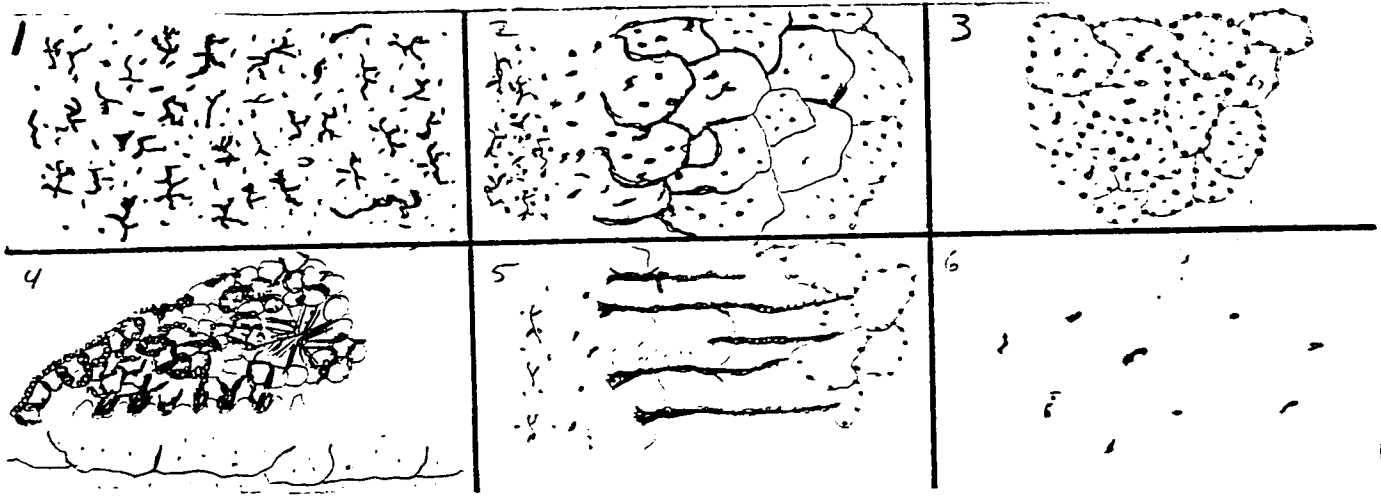
Figure 4 is a scanning electron micrograph of the as welded surface of the weld crown. The bead surface appearance is remarkably uniform. Figure 5 is an energy dispersive x-ray analysis of region A in Figure 5. Figure 6 is a corresponding analysis of region B. These analysis support the contention that region A is a cellular dendritic alpha region, and region B is a eutectic mixture. The volume fraction of the terminal liquid is significantly more than the the maximum predicted by solidification mechanics. Furthermore, the volume fraction of the eutectic constituent is greater in the weld toe and the partially melted regions, and slightly increased in the crown overlap region. In addition root passes examined showed enrichment in regions where a slight "U" groove was formed between the molten wall and the root pass crown. The largest effect was found in severely crowned beads. Figures 2 and 3 show a weld schematic indicating regions where enigmas are often seen, and indicating mechanisms proposed to account for these indications. These regions were produced using a technique developed by Nunes and Ding⁽²⁾. Several physical processes have been touted to account for the macrogregation (excess eutectic) noted in these cases, they are:

STRUCTURE -- RADIOGRAPHY CORRELATION

- (1) Fusion Zone
- (2) PMZ
- (3) Solid State
- (4) Toe Region
- (5) Overlap PMZ
- (6) Base Metal



X-ray, Plan View



Weld Transverse Cross Section

Shading represents X-ray intensity, microstructures drawn in insets.

Figure 3. Note the distribution of microstructural constituents. Sections 2 and 5 are each located in the PMZ. however region 5 shows back-filled cracks. These cracks are recognized by their size and orientation.

ORIGINAL PAGE IS
OF POOR QUALITY

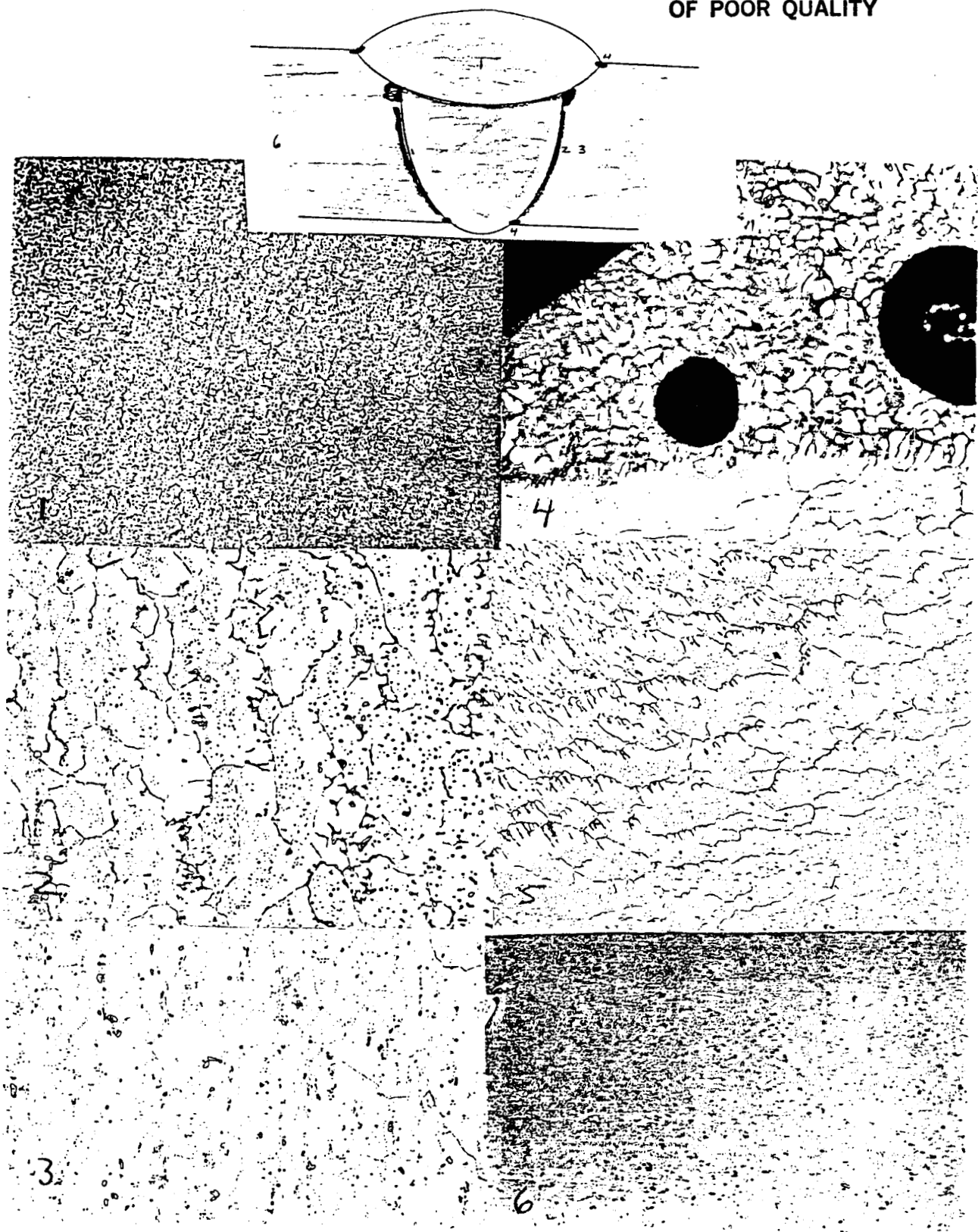


Figure 3a. Actual Microstructures Found in
The Regions Indicated

Inverse segregation
 Exudation
 Preferential Evaporation*
 Preferential Sputtering*
 Preferential Oxidation*
 Solid Bonding
 Divorced Eutectic Formation*
 High Frequency Effects*

These mechanisms marked by an asterik can also explain copper concentrations in excess of the eutectic observed by some investigators. Inverse Segregation: Occasionally, solute rejected during solidification is present at higher concentrations in solid regions that form during initial solidification. This can be explained on the basis of volume changes during solidification. Scheil⁽³⁾ generated an expression for the change in concentration produced by inverse segregation.

$$\text{Egn. 25} \quad \Delta C = C - C_o = \frac{M_{sL} C_{sL}}{A_e} + M_{kL} C_{kL} - C_o$$

(Assuming no pore formation)

C_o = nominal composition
 C = mean solute conc. at point
 M_{sL} = Mass of liquid/unit length at C_{sL} in liquid
 C_{sL} = Solute in liquid at eutectic temperature
 M_{kL} = Mass of solid unit length at C_{kL} in Solid
 C_{kL} = Solute conc. at solid interface at T_e
 A_e = Eutectic solid volume/eutectic liquid volume

Kirkaldy⁽⁴⁾ discussed the ramifications of this equation in rigorous mathematical detail. Relations between solute concentrations and solid and liquid densities must be used to evaluate this expression. The loss of solute in the liquid is equal to that gained by the solid (On the basis of total mass, the liquid loses solute, even though the concentration increases). However we must correct for that solute contained in fluid transported into the volume by volume contraction, and that solute transported out of the interdendritic region by contractions in the interdendritic channel. Interestingly, the specific volume of the solid is greater than the specific volume of the liquid for solidifying Al-Cu alloys over a wide range of temperatures. This does not provide a micromechanism for exudation, however, because the specific volume of the liquid decreases as a function of temperature at a rate in excess of the solid specific volume.

The extent of inverse segregation in structures solidifying under near equilibrium and under non-equilibrium conditions can be estimated by selecting appropriate values for the variables in equation 28. M_{ke} can be estimated as the volume fraction of the solid predicted at T_e , multiplied by the specific gravity of the mean solute concentration in the cored crystal. M_{se} can be estimated as the volume fraction of eutectic liquid when C_{se} is just reached, multiplied by the specific gravity of the liquid. These values can be determined from solidification mechanics in conjunction with the data of Pourier⁽⁵⁾ and Suaerwald⁽⁶⁾. Estimated values are shown in Table 2.

TABLE 2

	Equilibrium	Non-Equilibrium
M_{ke}	2.613	2.218
M_{se}	0.081	0.416
C_{ke}	0.0565% Cu	0.021% Cu
C_{se}	0.332% Cu	0.332% Cu
C_e	0.9411	0.9411

$$\Delta C \text{ equilibrium} = 0.002, (2\%)$$

$$\Delta C \text{ nonequilibrium} = 0.008, (8\%)$$

These numbers compare favorably with those predicted, and experimentally determined by Kirkaldy⁽⁵⁾. This segregation is not great enough to account for well defined enigmas. Simon and Jones⁽⁷⁾ noted that inverse segregation in Al-Cu alloys was more severe in reducing atmospheres.



Figure 4.

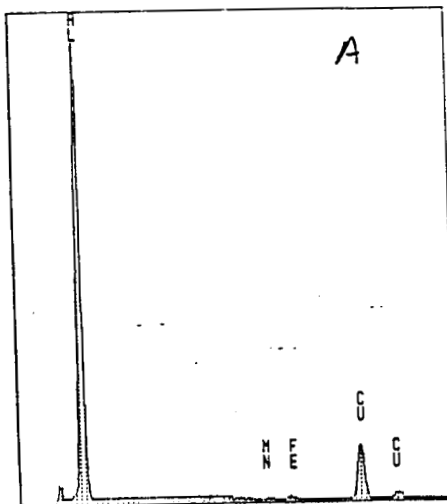


Figure 5.

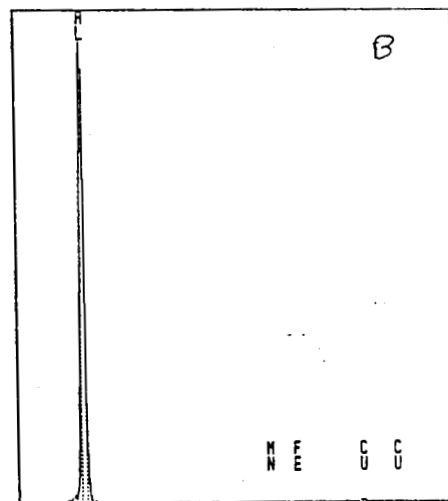


Figure 6.

The quantity D/R can be thought of as a "characteristic distance". At a distance of 5 (D/R) from the interface, less than 1% of the gradient remains. An expression for the "effective distribution coefficient," which takes not only a stagnant diffusive boundary layer, but a bulk mixing effect into account.

Egn. 24
$$K_E = \frac{K_0}{K_0 + (1-K_0) e^{-(R\delta/D)}}$$

where δ is the boundary layer thickness.

Note, as $\delta \uparrow$ $K_E \rightarrow K_0$

as $R \uparrow$ $K_E \rightarrow K_0$

as the rate of solidification increases the boundary layer thickness required to produce equilibrium solidification decreases.

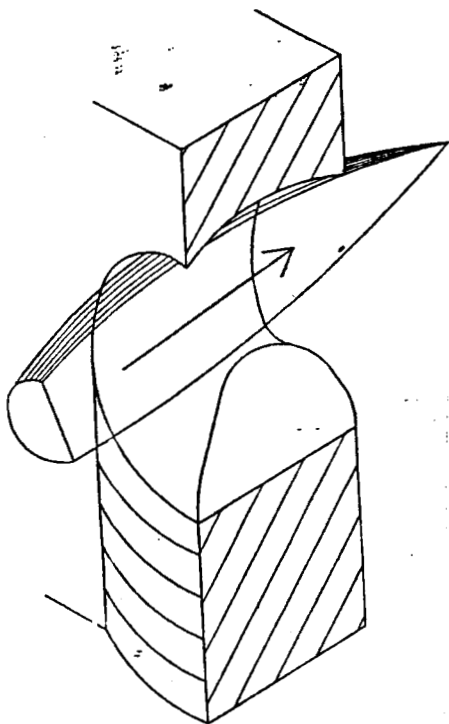


Figure 1. VPPA Keyhole

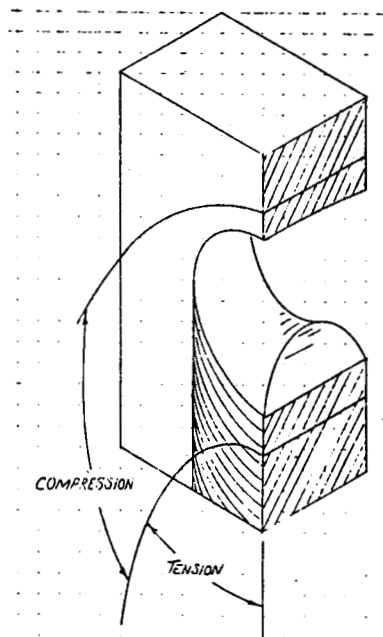


Figure 7. X-Y Stress around Keyhole (a). Exudate at Toe of Weld (b).



Segregation was most severe under atmospheres rich in hydrogen or water vapor. This segregation was related to the formation of gas bubbles as weld solidification progressed. Many small "bubbles" are noted in the toe regions of some welds in Al 2219. Figure 3a shows segregate enrichment near these bubbles. This effect will not produce the type of indication seen in Figure 3, but could lead to periodic x-ray intensity variation.

EXUDATION:

Residual stresses exist in all weld joints. However, the residual stress produced by welding does not influence fluid segregation. Transient stress, associated with the passage of the welding arc is more critical, and can effect solute distribution. Masubushi⁽⁸⁾ has demonstrated that compressive stresses exist in the weld region for a brief period before and immediately after the passage of the arc, Figure 7. The compressive stress in the weld region during solidification would promote exudation of enriched liquid from the intercellular regions. Figure 8 depicts the surface regions of a partially melted zone (pmz). Region A is a grain in the pmz, B a region of cellular solidification in prior weld metal and C the intercellular liquid. Elemental analysis in these regions are consistent with non-equilibrium solidification mechanisms proposed to account for the solute distribution and eutectic volume fraction. The intercellular region is typical of the alpha - CuAl₂ eutectic composition. Mn and Fe have also segregated to this region. Figure 3 depicts the typical structure of the eutectic. The arrow indicates a long needle, with a morphology typical of the more complex Cu-Al-Fe-Mn phase. The analysis of the solidified cell indicates a copper content lower than that found in the fluid encircled parent grain in the pmz. Cs in the cell can be estimated as 2.1% copper. That of the grain, subject to the thermomechanical history of the Al-2219 parent plate is near 5%. This is shown schemetically in Figure 14. The eutectic intercellular constituent is in relief at the weld surface, toe, and PMZ regions consistent with the compressive stress in these regions during welding.

Another effect based on a coupling of chemical segregation and mechanical stress is the enrichment noted in the "double" HAZ region. Nunes⁽⁹⁾ constructed a mathematical model of microfissuring in Inconel 718 welds. The model is based on an accumulated critical fracture strain. Interestingly, where Masubuchi relegates stress in the through plate (Z) direction to be zero in his transient stress calculation, Nunes assumes this to be the critical feature in microfissuring. This assumption was based on a study of crack morphology. The preponderance

of observed flaws indicate that the stress responsible for microfissuring is oriented perpendicular to the work piece. Figure 3 schematically depicts a common enigma found by transverse x-ray examination. This region is analogous, in two pass VPPA welds, to the hot-crack danger region found by Nunes. Figure 3 depicts this region in VPPA and GTAW welds in Al-2219. The weld procedure used promoted hot cracking in the base material. Copper enrichment in this region occurs by back filling of microfissures. Alloy 2219 is crack resistant because a large volume of terminal liquid can repair flaws that form. These regions are seen in the figure. A crack forms along the grain boundaries in the PMZ, and enriched liquid from the interface flows into the flaw by capillary action. Cracks are evident because of the directivity and extent of the fluid incursion. Interestingly, this leaves a depleted region adjacent to the flawed area, exactly what is noted on radiographs. Savage and Dix⁽¹⁰⁾ have described elemental distributions in back filled flaws.

Vaporization

Vapor is generated from the pool and surrounding environment during welding. The constitution of the vapor will not reflect that of the melt, higher vapor pressure components will be over-represented in the vapor⁽¹¹⁾. In a 2219 weld, the major constituent of the vapor is aluminum, whereas in a 2014 material Zn and Mg will be present. The minor alloy components have a great effect on the maximum pool temperature.

For a pure material

$$\text{Egn. 26} \quad \Delta G_e^\circ = -RT \ln P_a^\circ$$

by the Clausius-Clapeyron Equation

$$\text{Egn. 27} \quad \frac{dP}{dT} = \frac{S_v - S_l}{V_v - V_l}$$

or

$$\text{Egn. 28} \quad \frac{dP}{dT} = \frac{\Delta H_v}{T(V_v - V_l)}$$

therefore

$$\text{Egn. 29} \quad \log \frac{P_2}{P_1} = \frac{\Delta H_v (T_2 - T_1)}{2.303 R (T_2 - T_1)}$$

To extend the treatment to vapor loss from the solid, we need only to replace ΔH_v with ΔH_s , the heat of sublimation. Typically $\Delta H_s = (1.1) \Delta H_v$. For an alloyed material, the free energy change becomes

Egn. 30
$$\Delta G_e^\circ = \Delta \bar{G}_A$$

but

Egn. 31
$$\Delta \bar{G}_A = -RT \ln \bar{P}_A$$

$$\therefore \Delta G_{Tot}^\circ = RT \ln \bar{P}_A / p_a^\circ = RT \ln a_a$$

The relation can be used to calculate the equilibrium pressure of a species as a function of temperature. Dushman and Laferty⁽¹²⁾ and Matsuda⁽¹³⁾ have predicted the rate of material loss from the surface as

Egn. 32
$$V_a = 44.3 p_a (M/T)^{1/2}$$

for aluminum pools the maximum is approximately 10^{-4} g/scm². The maximum loss from a stationary fluid region at the weld toe is $\approx 10^{-5}$ g/SCM², and that in the solid regions adjacent to the toe is $\approx 10^{-7}$ g/SCM². A 1 cm² surface of aluminum in the partially melted region will lose

$$(10^{-7} \text{ g/scm}^2) (27 \text{ g/mole})^{-1} \left(\frac{6.02 \times 10^{23} \text{ atoms}}{\text{mole}} \right)$$

or

$$\approx 10^{15} \text{ atoms/scm}^2$$

Because the surface exposes 6.25×10^{14} atoms/cm², it will lose 2 to 10 layers per second. The surface is exposed to temperatures in the range discussed for 5 seconds. Thus it will lose 10 - 50 atom layers of aluminum. This is enough enrichment to be seen in Edax or Auger examination. Further surface enrichment is caused by sputtering. When a neutral atom or a charged species strikes a surface and causes a neutral atom to be removed for that surface the surface is said to be sputtered. Pattee⁽¹⁴⁾ et al have shown that this occurs only under the arc region during reverse polarity segments in plasma welding. Argon atoms sputter

2219 efficiently, but Helium does not. Therefore Argon is used as the plasma gas because it removes arc-disrupting, rectifying layers of aluminum oxide from the work surface. Sputtering action removes aluminum with greater efficiency than it removes copper, owing to the mass of target atoms. The greater affinity of aluminum for oxygen will also promote a copper enrichment at the surface. Stationary VPPA welds made on a vertical surface were examined by optical and scanning electron microscopy. Large volumes of eutectic liquid have formed on the surface of the bead. These are accumulated volumes of eutectic liquid. Elemental analysis in these regions did occasionally indicate concentrations of copper in excess of the eutectic.

Several authors⁽¹⁵⁾ have advanced solute banding, promoted by velocity changes in the pool, as a mechanism for weld toe solute enrichment. Though it is true that rich solute bands are formed, and that their frequency is greater at the pool edge, so too are solute poor bands formed and their frequency is greater at the edge. There is no net effect on solute produced.

In summary, enigmas are produced by the action of microsegregations mechanisms intimately related to all welding processes. When coupled with the mechanical stress state near the pool, these mechanisms can produce exaggerated concentration of copper in well defined areas. Further concentration of copper in crown toe area can occur by evaporation, sputtering, and oxidation. Root toe areas are formed by a similar mechanism, but no further enrichment by sputtering can occur.

Arc Characterization

The measured amperage and voltage, for the range of conditions studied is shown in Table 1. In aluminum a rise in voltage of $\sim 8V$ is encountered on switching to reverse polarity, Figure 9. In both straight and reverse polarity the voltage increases as standoff increased. The amperage associated with the straight polarity portion of the cycle increased so as to provide a time averaged current that approximates the program current. It does not, however, average power.



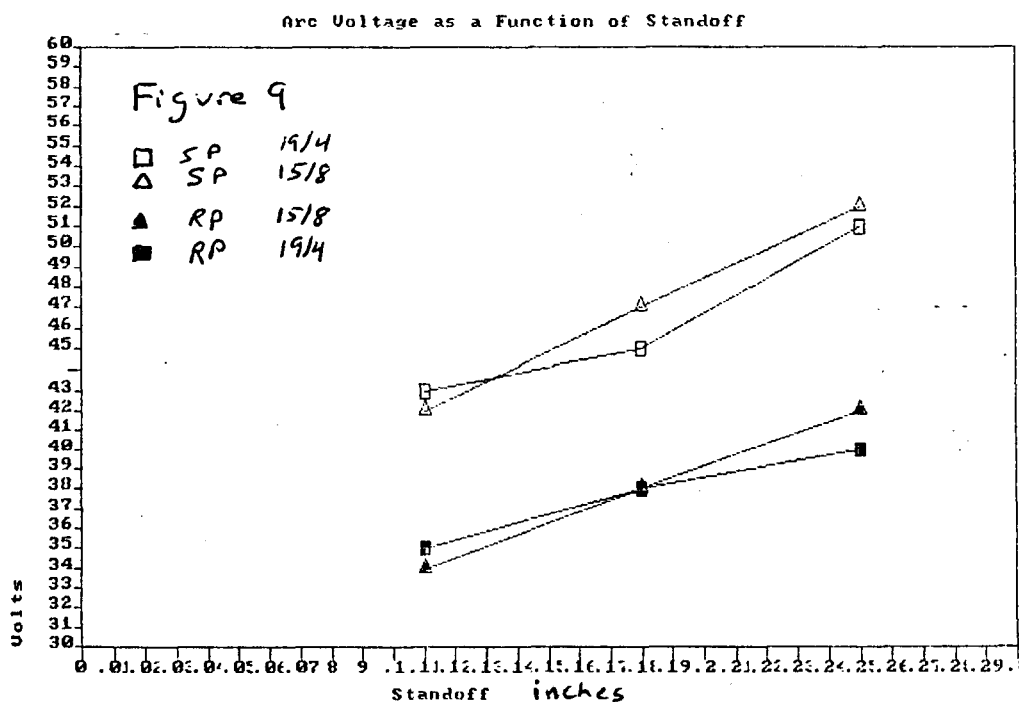
Figure 8. The Partially Melted Zone (PMZ).

Table 1.

ORIGINAL PAGE IS
OF POOR QUALITY

Plate	Amperage		Voltage		Polarity	Standoff	Gas Flow	Current
	S	R	S	R				
M1	270	225	38	45	19/4	0.180	8.1	260
M3	275	220	42	51	19/4	0.250	8.1	260
M5	270	225	35	43	19/4	0.110	8.1	260
M7	280	225	38	53	19/4	0.180	10.1	260
M9	290	230	37	51	19/4	0.180	5.1	260
M11	300	215	38	47	15/8	0.180	8.1	260
M13	240	185	38	46	15/8	0.180	8.1	210
M15	330	235	45	52	15/8	0.180	8.1	290
M17	280	210	43	52	15/8	0.250	8.1	260
M19	275	230	34	43	15/8	0.110	8.1	260
M23	315	230	37	47	12/12	0.180	8.1	260
M27	270	225	34	42	19/4	0.180	8.1	260

R= reverse cycle S= straight cycle current in amps, potential in volts, standoff in inches, gas flow in cfh.



APPENDIX I

Historical Perspective - Alloy 2219

Aluminum alloy 2219 was initially examined as a candidate for spacecraft applications in March of 1962; during the SATURN program. Designers were well aware that the performance of high strength aluminum structures were limited by the characteristics of the weld joints in these materials. Despite their inherent problems, Al-Cu alloys were desirable because of favorable strength to weight ratios and the lack of a low temperature ductility transition. The former attribute provides weight savings in structural applications, the latter makes Al-2219 the aluminum alloy with the greatest range of service temperature. Aluminum 2219 is basically a binary Al-Cu, age hardening material. Its predecessors, 2014 and 2024 are Al-Cu materials with substantial Mg and Zn added to enhance aging kinetics at low temperatures. Alloy 2219, heat treated in equivalent fashion is not as strong as 2024 or 2014 at low temperatures, but has superior high temperature strength as well as superior weldability. Improved weldability is won by increased copper content in the alloy. Compositions of Alloy 2014, 2024 and 2219 are shown in Table A1. Enhanced weldability is caused by an increased volume fraction of liquid during critical periods of solidification. The increase in high-temperature strength is caused by three factors:

1) Alloy 2219 is age hardened with a precipitates. The kinetics of formation of this phase are sluggish vis-a-vis the S' and S Al-Cu-Mg precipitates that strengthen 2014 and 2024. Mg reacts strongly with vacancies and accelerates nucleation of precipitates. S' forms rapidly during natural aging and overages quickly at elevated temperatures.

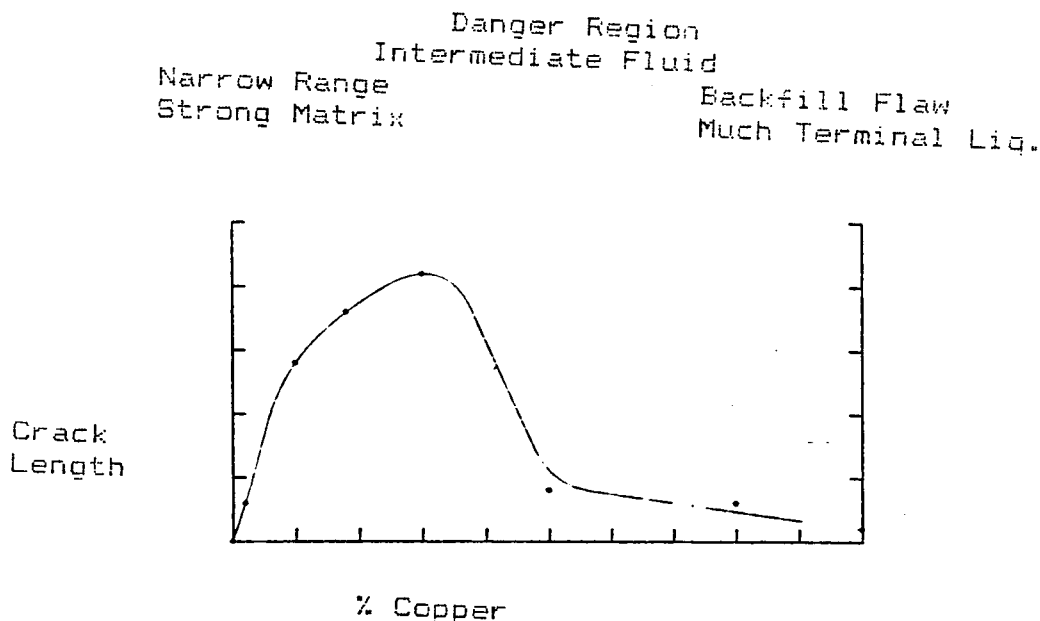


FIGURE A1

2) Mn, Fe, Zr, Ti, and V are present in Alloy 2219. These elements, though present at low concentrations, raise the re-crystallization temperature markedly, and help to pin grain boundaries.

3) The increased fraction of second phase constituents at the grain boundary, caused by increased copper, help prevent the onset of grain boundary sliding.

Aluminum alloy 2219 is used to fabricate the Space Shuttle Main Tank primarily because of weldability considerations. Further, a large data base of cryogenic applications has accumulated since 1962. Alloy 2219 has been used in aircraft parts and structures subject to elevated temperatures as well, therefore the mechanical properties are well understood.

Weldability of Al-Cu Alloys

Two major weldability concerns, with precipitation hardened aluminum alloys, are their susceptibility to hot cracking and their drastic reduction in weld composite region strength. Jennings and Pumphrey⁽¹⁶⁾ did extensive empirical work treating the weldability of binary aluminum alloys. This work was critical to Borlands⁽¹⁷⁾ classic discussion of hot cracking. In all cases hot cracking is predicated on a stress developed across liquid bridges between adjacent solid regions. These stresses can be produced by solidification shrinkage of weld metal in concert with thermal contraction of the workpiece. Material with a continuous liquid network along grain boundaries is still capable of supporting loads. Grotke⁽¹⁸⁾ has shown that the strength is related to the film thickness.

Egn A1

$$\sigma = \frac{2\gamma_L}{h}$$

Even with a continuous liquid 750°A thick in the boundry, the material can support a load of 5 ksi. The hot cracking sensitivity of aluminum alloys is critically dependent on the weld metal chemistry. Pumphry⁽¹⁹⁾ has shown that the crack sensitivity of an Al-Cu alloy increases to a maximum near a nominal Cu content of just less than 4% Cu, and then falls to a minimum level near 6% Cu, Figure A1. Borland explains this behavior by postulating a temperature range during solidification during which there is a coherent, load carrying, cellular matrix woven through a liquid that remains to solidify. The material is of low strength and minimal ductility in the temperature range coherency and the solidus. The wider the temperature range between coherency and the final solidification, the greater the tendency to hot crack, because total strain will be directly related to the solidification interval.

If sufficient liquid is present during the latter stages of solidification, incipient cracks that form may be refilled by capillary action. The liquid eutectic, because it freezes at a constant temperature, is the most effective regenerating fluid. Conversely, a dearth of eutectic may aggravate cracking by creating zones of weakness in an interdendritic film. Jennings⁽¹⁶⁾, postulated that maximum cracking occurs at an alloy composition with the greatest range between the solidus on the liquids (the maximum solid solubility). However, the solidification of welds is a non-equilibrium process. Constitutional supercooling greatly changes the volume of terminal eutectic phase produced, and depresses the coherence temperature. Maximum weld cracking occurs at lower copper concentrations than predicted for equilibrium.

To summarize:

1) At low Cu levels, the amount of eutectic terminal transient is not sufficient to form a continuous grain-boundary network. The fraction solidified is large even at slight undercoolings, and the solid forms a strong, crack-resistant network.

2) At intermediate Cu concentrations, the volume of eutectic fluid becomes just sufficient to form a continuous interdendritic liquid, and the crack sensitivity is large.

3) At Cu concentrations near that found in alloy 2219 crack resistance is large. Because the solidification range is diminished, minimizing shrinkage strains, and because the volume fraction of eutectic is sufficient to "back-fill" cracks as they form.

Table A1. Composition of Common Al Alloys

Alloy	Al	Cu	Mn	Mg	Si	V	Zr	Ti
2219	93	6.3	0.3	-	0.8	0.1	0.2	0.1
2014	93.5	4.4	0.8	0.5	-	-	-	-
2024	93.5	4.4	0.6	1.5	-	-	-	-
5456	93.9	-	0.8	5.1	-	(0.12 Cr)	-	-

Conclusions
and
Recommendations

- 0 Alloy 2219 is a uniquely weldable binary alloy of aluminum and copper. It is weldable because large volumes of liquid are available during its latter stages of solidification.
- 0 Transient stresses in the weld region produce macrosegregation by causing exudation and by causing flaws that are subsequently back filled. These are the major mechanisms for producing macrosegregation in Al-2219 welds.
- 0 Apparent enrichment, beyond the eutectic composition was found at the weld surface. Sputtering and evaporation can account for this increase.
- 0 Radiographic contrast effects, including the straight line indication (but not diffraction effects) are a result of transient stress induced macrosegregation.
- 0 Contrary to conventional wisdom, more heat is delivered to aluminum, steel, and inconel materials during the reverse polarity portion of the VPPA welding cycle than during the straight polarity segment.
- 0 Current in the welding arc falls during the reverse polarity portion of the welding cycle, but voltage increases. The voltage increases so much that greater power is often produced in the reverse cycle.
- 0 At any programmed current the current in the straight polarity portion of the cycle increases as the proportion of reverse polarity increases.

REFERENCES

- 1) Walsh, D.W. Summer Report "Physics of the VPPA Weld Keyhole Region; Pool Flows" NASA-ASEE Summer Faculty Program, August 1987
- 2) Nunes, A.C. Jr., and Ding, J., Private Discussions, Pending Publication, July 1988.
- 3) Schiel, E., Metallforschung, 1947, V2, p69.
- 4) Kirkaldy, J.S. and Youlodos, W.V., "Contributions to the Theory of Inverse Segregation" Met. Trans. 212, Dec. 1958.
- 5) Pourie, and Ganesco, "Densities of Al Rich Al-Cu Alloys During Solidification" Met. Trans. A 18 Apr. 1988
- 6) Suarerwals, F. Metallwirtschaft, 1943, V22, p543.
- 7) Adams, D.E., "Segregation in Al-Cu Alloys" J. Inst of Metals 75 1949 p809
- 8) Masubuchi, K. and Simmons, F. "Analysis of Thermal Stress and Metal Movement During Welding" Redstone Rep. RSIC-820 July 1968.
- 9) Nunes, AC, "Microfissuring in IN718" NASA TM-82531 June 1983.
- 10) Savage and Dickinson, "Electron Microanalysis of Backfilled Hot Cracks in Inconel 600. WJ 11, 1972, p543.
- 11) Block-Boulton and Eagar, "Selective Evaporation of Metal From Weld Pools" Trends in Welding Research, Conf. Proc. ASM 1981
- 12) Dushman and Laferty, Scientific Foundations of Vacuum Technique John Wiley 1962 p696.
- 13) Matsuda and Hashimoto, "Penetration Mechanisms in EB Welds" Trans. Nat. Res. Inst. for Metals (Jap.) V7, 5 1965
- 14) Pattee Meister and Monroe, "Cathodic Cleaning and Plasma Arc Welding of Aluminum" Welding Journal, May 1968.
- 15) D' Anessa, M.J., "Solute Redistribution in the Weld Pool" Welding Journal, May 1966.
- 16) Jennings, Singer, and Pumphrey, Journal of the Institute of Metals, 74, 227, 1948 "Hot Shortness of Some High Purity Alloys in the Al-Cu System"
- 17) Borland, J., British WJ 7, 508, 1960. "Supersolidus Cracking"
- 18) Grotke, P., "Physical Metallurgy of Hot-Ductility Testing" Welding Journal, Nov. 1970.
- 19) Pumphrey and Lyons, J. Institute of Metals, 74, 439, 1948 "Cracking During the Welding of the More Common Binary Aluminum Alloys".

N89-21756

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

THE ROLE OF THE TURBULENT PRANDTL NUMBER
IN TURBINE BLADE HEAT TRANSFER PREDICTION

Prepared by: Kevin W. Whitaker
Academic Rank: Assistant Professor
University and Department: The University of Alabama
Aerospace Engineering

NASA/MSFC:
Laboratory: Structures and Dynamics
Division: Aerophysics
Branch: Computational Fluid Dynamics

MSFC Colleague: Helen V. McConnaughey

Date: August 4, 1988

Contract No.: NGT 01-002-099
The University of Alabama

XXX

THE ROLE OF THE TURBULENT PRANDTL NUMBER
IN TURBINE BLADE HEAT TRANSFER PREDICTION

by

Kevin W. Whitaker
Assistant Professor of Aerospace Engineering
The University of Alabama
Tuscaloosa, Alabama

ABSTRACT

A study was undertaken to improve the prediction of external (gas-to-blade) heat transfer coefficients in gas turbine engines. The study specifically investigated the effects of improved eddy diffusivity of heat modeling in the turbulence model. A two-dimensional boundary layer code, STAN5, was selected and modified by incorporating several different turbulent Prandtl number models. Results indicated that slight effects were attributable to the modified turbulence model. Boundary layer character appeared to be much more significant.

ACKNOWLEDGEMENTS

I would like to thank the National Aeronautics and Space Administration along with the American Society for Engineering Education for sponsoring this fabulous opportunity. I have found the Summer Faculty Fellowship Program to be extremely beneficial and enjoyable. Without question, the program has played an important role in my professional development.

Sincere appreciation is due Helen V. McConnaughey whose interest in the Summer Faculty Fellowship program provided me with what turned out to be a very interesting and timely project. I also appreciate the valuable assistance provided by L. Michael Freeman and Ernestine Cothran throughout the course of the program.

INTRODUCTION

The thermal aspects of blade design is one of the more difficult engineering tasks facing a designer of any modern gas turbine engine. Thermal (and many times aerodynamic) analysis procedures currently available to designers have deficiencies that do not permit achievement of design goals without expensive experimental development programs. For example, the external (gas-to-blade) heat transfer coefficient still eludes satisfactory prediction using computational fluid dynamic codes. Even if consideration is restricted to the nominally two-dimensional midspan region of a turbine blade, prediction is still unsatisfactory. The reasons for the unsatisfactory prediction capability of the codes are complex but ultimately lie in the fundamental concepts and models used to define the fluid dynamic and heat transfer behavior. Without question, the complex gas turbine engine environment pushes current models to their limit. Thus, there exists a need for an improved design approach making use of codes with sufficiently improved turbulence modeling.

The work presented here was undertaken to improve the prediction of gas-to-blade heat transfer coefficients. Specifically, it investigates the effect of modeling the eddy diffusivity of heat via several turbulent Prandtl number models published in the literature.

OBJECTIVE

The objective of this study was to improve the computational prediction of the external (gas-to-blade) heat transfer coefficient for gas turbine engine applications. Such an improvement would reduce and perhaps eliminate the expensive experimental iterations that current engine designers must endure. The end result would impact engine design in a very positive way.

PROCEDURE

CODE SELECTION

Current gas turbine engine design practice is to use a two-dimensional boundary layer analysis to calculate the gas-to-blade heat transfer coefficients. Certainly any computational method which does not solve the full time-dependent Navier-Stokes and energy equations cannot be expected to be universally valid over the entire range of circumstances governed by these equations. However, there are solutions from reduced sets of these equations that are valid for a subset of problems. Such is the case here where it is implied that the flow field immediately adjacent to the surface of an airfoil in typical gas turbine geometries can be analytically modeled using boundary layer equations.

Perhaps the most familiar and widely used boundary layer method is a finite difference technique which relies on algebraic relations for defining turbulence quantities. A very common design tool of this type is STAN5, a code developed by Crawford and Kays [1] and later modified by NASA Lewis Research Center [2]. For boundary layer flow with heat transfer, STAN5 involves the solution of two governing partial differential equations using the numerical scheme of Patankar and Spalding [3]. Turbulence closure is obtained using eddy diffusivity concepts. The STAN5 code has received wide attention because of its careful development, flexibility, and adequate documentation. For those very reasons, STAN5 was selected to be used for this study.

The STAN5 code allows many parameters to be adjusted and it was felt that one set of parameters should be selected and held constant throughout the test so that the influence of the turbulent Prandtl number models could be determined. Of course it was desirable to have the parameters describe a true gas turbine engine flow field as closely as possible.

Reviewing published data for flow over turbine blades, it was decided that a fully turbulent boundary layer on both the suction and pressure surfaces of the blade would be assumed. This is perhaps a point of contention but it was adopted for a couple of reasons. First, many transition models have been tried in the past with limited success [4]. Secondly, a typical gas turbine engine environment flow field has a high free stream turbulence level. Also, any

boundary layer character change (such as relaminarization) that might occur would be modeled through the pressure gradient implicitly contained in the input data.

STAN5 has two eddy diffusivity models, the Prandtl mixing length hypothesis (MLH) and the higher order turbulent kinetic energy (TKE) concept. For this study, the MLH method was selected based on the past attention given to it - especially in gas turbine engine studies. Also the choice of the MLH model can be considered a practical selection. The detailed experimental data required to realistically tune higher order turbulence models for gas turbine engine applications are quite scarce. On the other hand, the global-type boundary layer data normally used to develop lower order turbulence models (such as the MLH) are more common.

Another consideration was whether to assume the blade surface was a flat plate or to include the blade curvature into the analysis. A curvature model was available in STAN5 but previous studies [4] have revealed that using the curvature model did not significantly effect the heat transfer results. Also, as pointed out earlier, current design practice is to assume the flat plate. Therefore a flat plate model of the blade was assumed in this study.

Finally, all specifiable constants in STAN5 were set equal to values suggested by Crawford and Kays.

TURBULENT PRANDTL NUMBER MODELS

As part of the eddy diffusivity concept used in the STAN5 turbulence model, a parameter called the turbulent Prandtl number (Pr_t) is introduced. This dimensionless parameter links the eddy diffusivity of momentum (ϵ_m) and the eddy diffusivity of heat (ϵ_h). The turbulent Prandtl number concept mirrors the classical laminar approach where the momentum and thermal transport mechanisms are related by the molecular Prandtl number (Pr). By definition, the turbulent Prandtl number is:

$$Pr_t = \epsilon_m / \epsilon_h .$$

Typically, ϵ_m is solved for using a mixing length hypothesis or the turbulent kinetic energy concept. Then ϵ_h is determined assuming a $Pr_t = 1.0$ (Reynolds Analogy) or some other constant value ($0.8 \leq Pr_t \leq 0.9$ has received wide acceptance for gas flows). This whole premise of using a constant Pr_t totally ignores the heat transport mechanism. Direct modeling of ϵ_h would provide a much more realistic

picture of the flow physics. The individual ϵ_m and ϵ_h models could then be combined to yield a Pr_t model which contains much more information.

A number of attempts have been made to predict turbulent Prandtl numbers containing ϵ_h modeling through highly idealized analyses. Many of the resultant models are fundamentally based on an idea first suggested by Jenkins [5]. He hypothesized that a turbulent eddy, while moving transverse to the mean direction of flow, may lose heat at a different rate than it loses momentum. His analysis assumed that the eddy lost heat by simple molecular conduction and lost momentum by the action of viscous shear. He also modeled the eddy as a spherical element of fluid with a radius equal to the size of the mixing length.

Turbulent Prandtl number models for this study were obtained after an extensive review of the literature. To establish a baseline, a constant turbulent Prandtl number model using $Pr_t = 0.86$ was used. Then four different models were selected in order to test the validity of the various assumptions contained in them. Each of the models is briefly described in the following. For a complete description of each method the reader is referred to the specified reference.

0 Crawford and Kays [1]

Reflecting on Jenkin's hypothesis, Crawford and Kays suggested that the turbulent eddies transfer momentum by the action of impact and pressure forces and that viscous forces are not involved. The success of the mixing-length theory in which viscosity is not a variable would suggest this to be the case. Conversely, there is no mechanism other than molecular conduction whereby heat can be transferred from an eddy. This means that the transport mechanisms must be different. Their final turbulent number model is in terms of the molecular Prandtl number and the eddy diffusivity of momentum.

0 Thomas [6]

Feeling that a fresh approach was needed, Thomas developed a turbulent transport model employing an elementary surface renewal and penetration model. He based his model on the idea of diffusive penetration of eddies through a film, (in this case the viscous sublayer), intermittently renewed by fluid from the region of turbulent flow. He further assumed that the molecular transport is predominate during the time the fluid elements are in the vicinity of the surface. The analogy between heat and

momentum transfer is a result of the renewal mechanism and the resultant model is a function of molecular Prandtl number, eddy diffusivity of momentum and a normalized distance from the surface (y^+).

0 Cebeci [7]

Cebeci developed a model for the turbulent Prandtl number based on Stokes flow considerations. Expressions for both ϵ_m and ϵ_h were obtained by following Van Driest's damped mixing-length representation of the Stokes flow viscous sublayer. It differs from the other models in that his ϵ_h expression provides a continuous temperature distribution across the boundary layer and also accounts for any pressure gradient. The model expresses the turbulent Prandtl number in terms of the molecular Prandtl number and a normalized distance from the surface (y^+).

0 Tyldesley and Silver [8]

The approach taken by Tyldesley and Silver is quite different from the approaches discussed previously. They abandoned the mixing-length concept and investigated the transport properties of a turbulent fluid by using a simple model to represent the detailed fluid behavior. In the model fluid behavior is attributed to the motions of fluid entities of varying size, shape, and velocity. Their analysis enables them to find expressions for the eddy coefficients of momentum and heat in terms of properties of the turbulence. For example, their model indicates that the turbulent Prandtl number is a function of not only molecular Prandtl number and Reynolds number but also turbulence intensity as well.

EXPERIMENTAL DATA

In order to evaluate the predictive capabilities of any computational method, it needs to be compared with experimental data. Many well documented heat transfer studies have been performed and there is a fair amount of reliable data available. This study used the work performed at Detroit Diesel Allison by Hylton et al. [4]. The main reason for selecting this data was that in addition to presenting their experimental results, the authors also provided the necessary STAN5 input data for their experimental configuration. This eliminated the need to develop the required input data thus allowing more time to be devoted to the task at hand.

The experimental program of Hylton et al. studied flow

through a turbine cascade. The cascade contained three blades that were characteristic of a first-stage turbine. The blades were designated as "C3X" airfoils and the profile of one is shown in Figure 1. The center blade in the cascade was instrumented and provided the aerodynamic and heat transfer data. The operating conditions for the data set used for comparison in this study are given below.

Inlet Total Temperature:	1460°F
Inlet Mach Number:	0.16
Inlet Reynolds Number:	640,000
Free-stream Turbulence Level:	6.55%
Blade Surface Temperature:	1182°F

Note: The inlet Reynolds number is based on true chord.

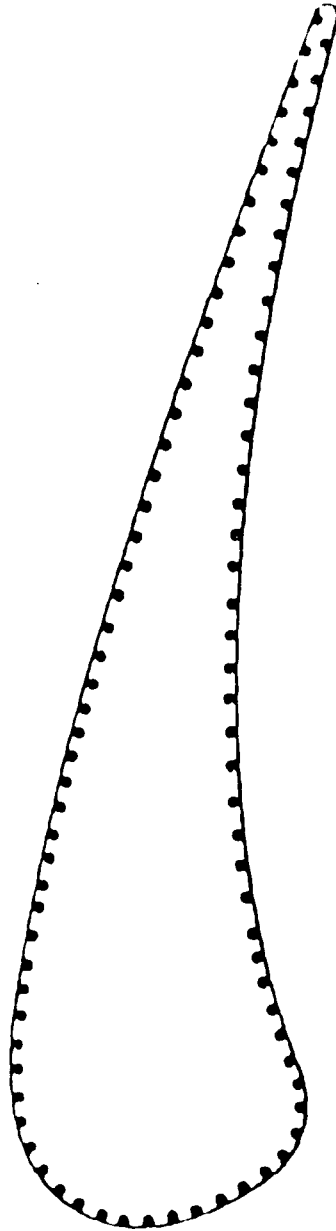


Figure 1. Profile of C3X Turbine Blade

RESULTS

The heat transfer coefficient predictions produced by STAN5 combined with the various turbulent Prandtl number models can be seen in Figures 2 and 3. For presentation, the heat transfer coefficient (H) has been normalized by a reference value (H_0) of 200 BTU/Hr/ft²/°F and the distance along the blade surface (S) is normalized by the total surface arc length (ARC). Also shown with the predictions is the experimental data of Hylton et al.

Figure 2 shows the distribution of the heat transfer coefficient on the blade's suction surface. It can be seen that for the first 20 percent of the surface the models yield identical predictions. A large favorable pressure gradient exists in this region and the predictions are likely representative of a boundary layer forced to relaminarize. This is confirmed by observing the laminar solution which is also shown on the figure. At a surface distance of about 20 percent the flow appears to transition. All the models predict the transition start but then have limited success downstream. The constant Pr_t assumption and the model of Cebeci overpredict the transition but recover nicely downstream. The model of Thomas predicts transition very well but then immediately underpredicts the heat transfer coefficient along the rest of the blade surface. The Crawford and Kays model yields results similar to Thomas' model but does not underpredict as severely downstream.

Distributions on the blade's pressure surface are shown in Figure 3. It is obvious that none of the turbulent Prandtl number models adequately predict the distribution represented by the experimental data. The laminar solution is also shown for comparison and reveals that the turbulent Prandtl number predictions do not vary from the laminar solution until a surface distance of 60 percent. After that the models predict what appears to be a transition from laminar to turbulent flow with a corresponding increase in heat transfer coefficient.

The turbulent Prandtl number models did not appear to significantly alter the heat transfer predictions and at times, there was little difference between the laminar and turbulent flow assumptions. This suggests that the driving force behind this phenomenon is the character of the boundary layer. One of the assumptions made in this study was that of treating the airfoil as a flat plate and using

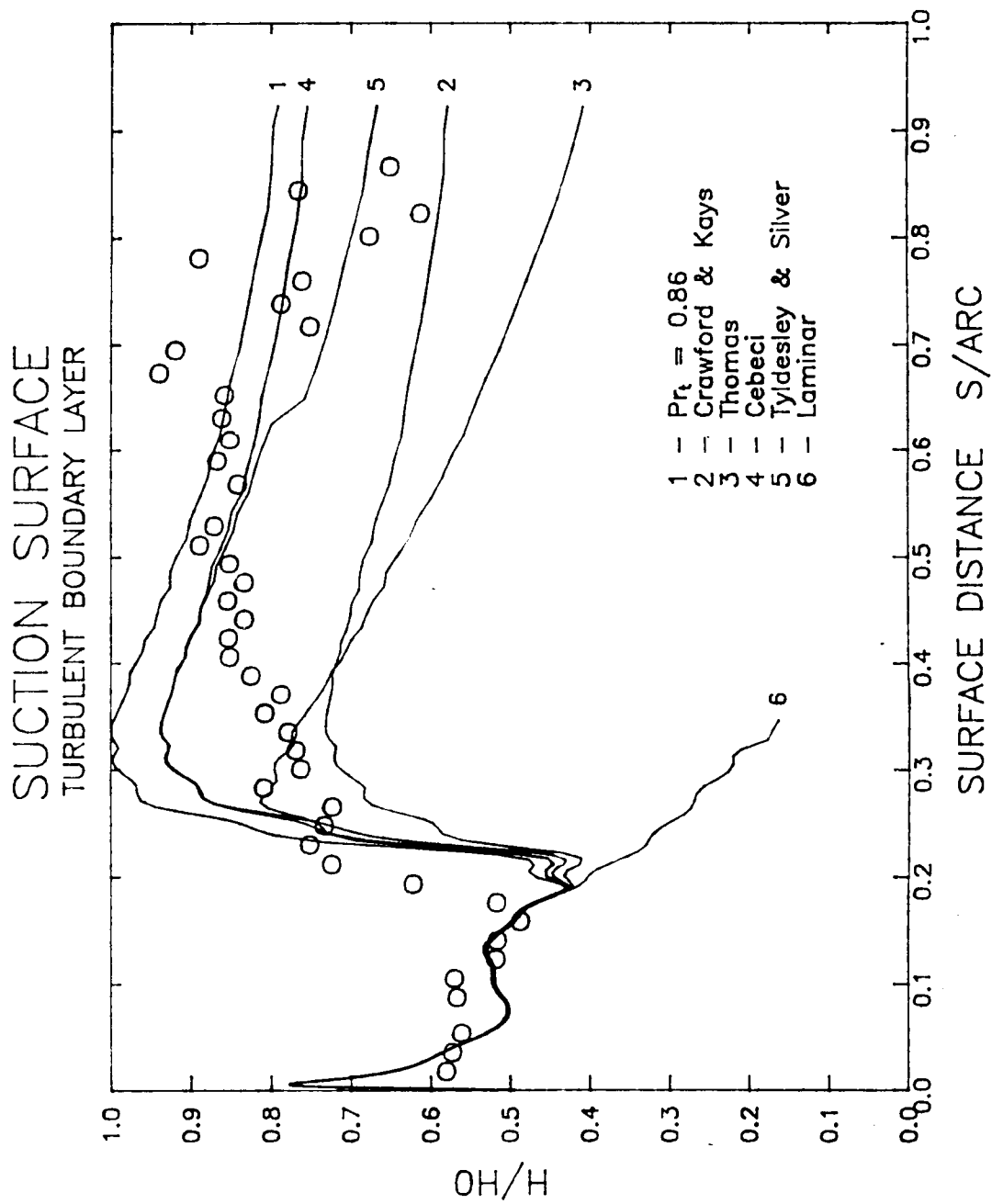


Figure 2. Heat Transfer Coefficient Distribution on the Suction Surface

PRESSURE SURFACE TURBULENT BOUNDARY LAYER

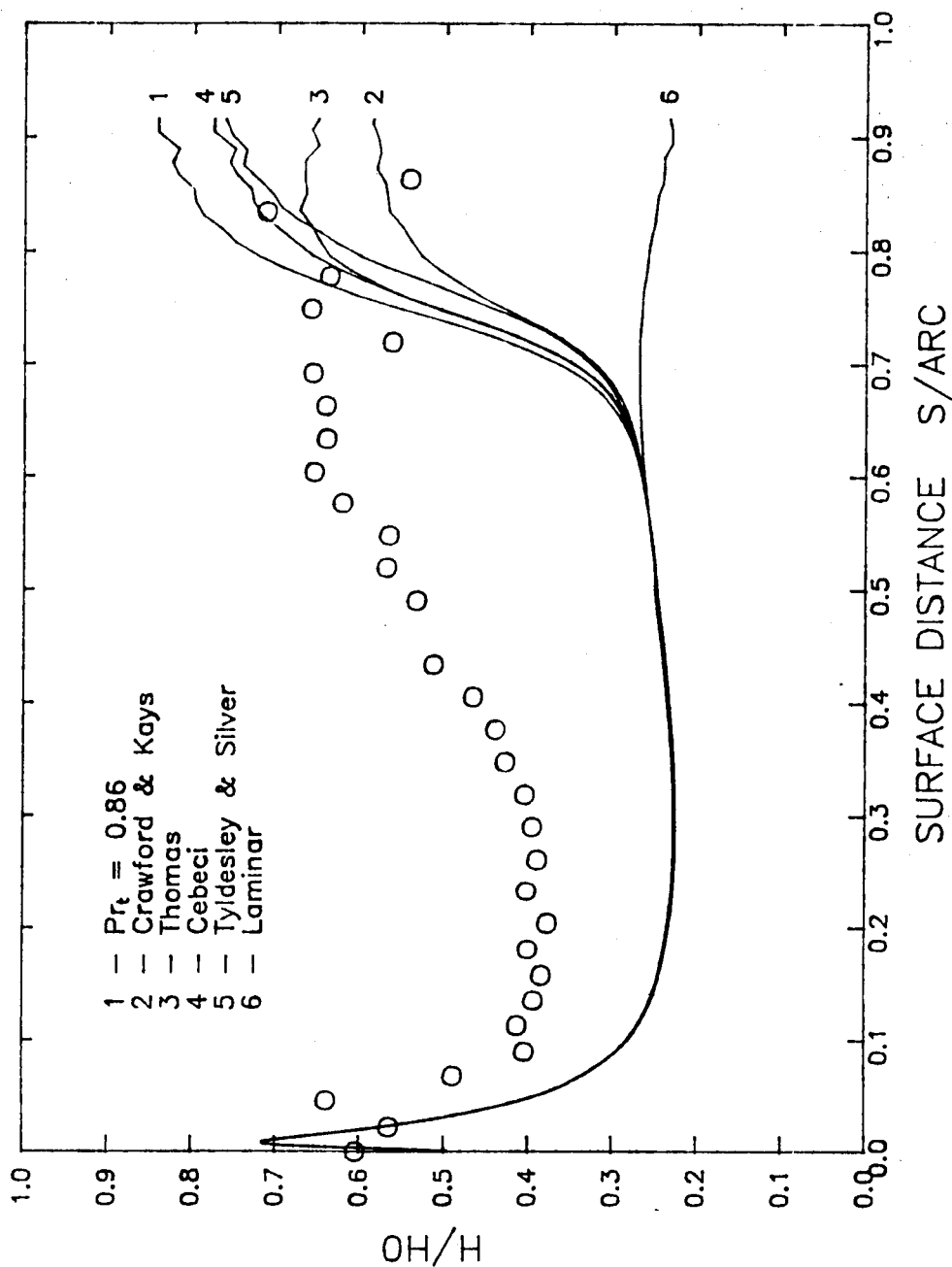


Figure 3. Heat Transfer Coefficient Distribution on the Pressure Surface

all of the correlations accompanying that assumption. Within STAN5 there is an expression to determine the effective viscous sublayer thickness (A^+) which is based on flat plate analyses. Also, there is a correction for A^+ depending on the pressure gradient experienced by the flow. An attempt was made to see how well these concepts apply to turbine blades.

After some "trail-and-error", it was found that the A^+ correction factor was significantly effecting the heat transfer coefficient predictions. This can be seen in Figure 4. On the pressure surface very good predictions were obtained with all the models by not using the A^+ correction factor. Thomas' model predicts the trend exhibited by the experimental data very well. The model of Crawford and Kays also agrees well but consistently underpredicts. The other models predict well up to a surface distance of about 60 percent and then overpredict for the remainder of the blade. The suction surface predictions did not respond in a similar manner, however. In fact, not using the A^+ correction factor caused the predictions to become even worse. An example of a typical result is shown in Figure 5.

An attempt was made to further investigate the character of the suction surface boundary layer. Clearly both laminar or fully turbulent flow do not represent what is happening on the suction surface. One set of predictions were obtained assuming that the boundary layer on the suction surface was in transition from leading edge to trailing edge. Representative results are shown in Figure 6. Although the transition seen previously at 20 percent still exists, it is not as abrupt and the downstream levels are not correct. Examples with and without the A^+ correction factor are shown.

PRESSURE SURFACE TURBULENT BOUNDARY LAYER

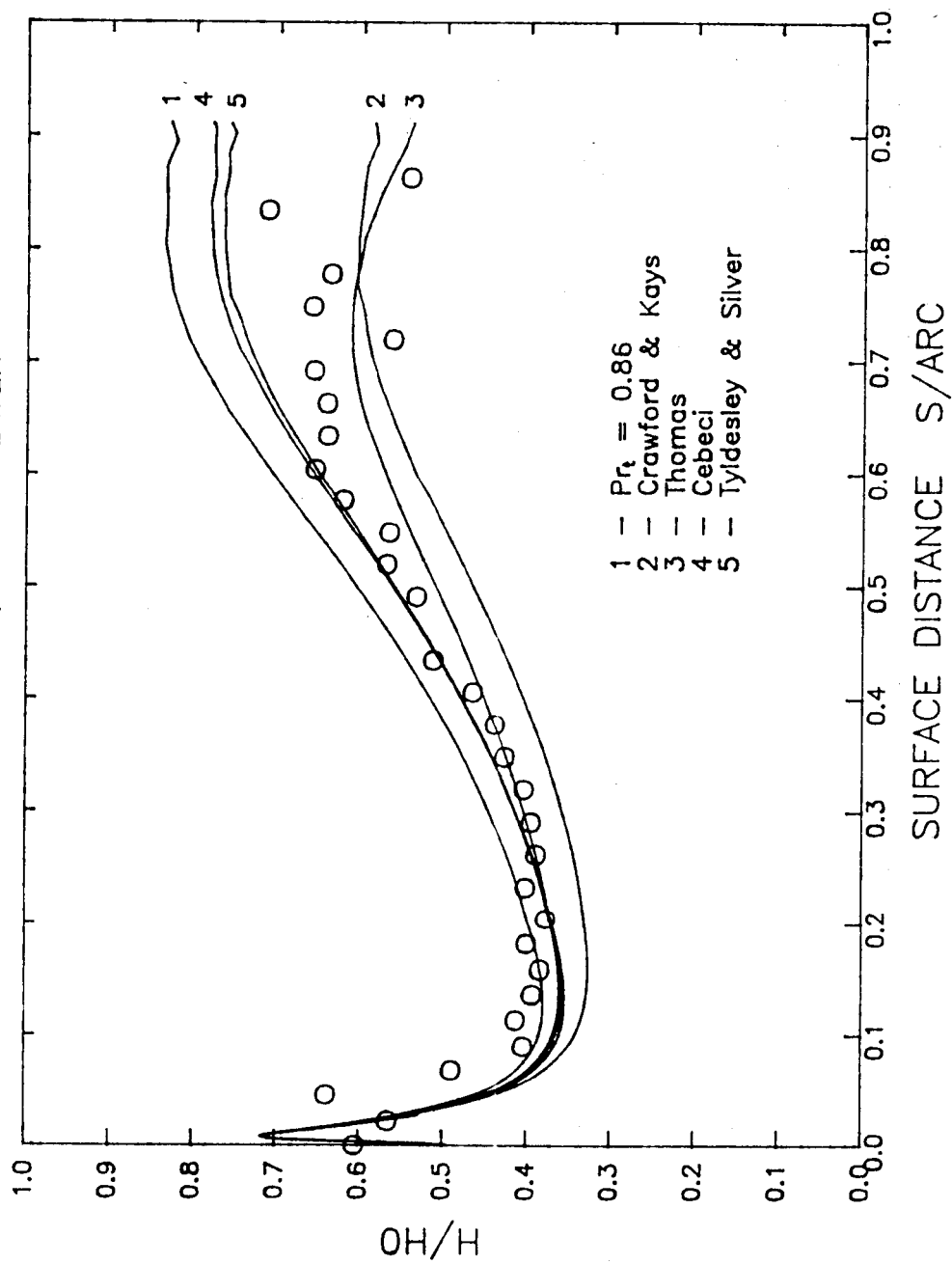


Figure 4. Heat Transfer Coefficient Distribution on the Pressure Surface without using A+ Correction

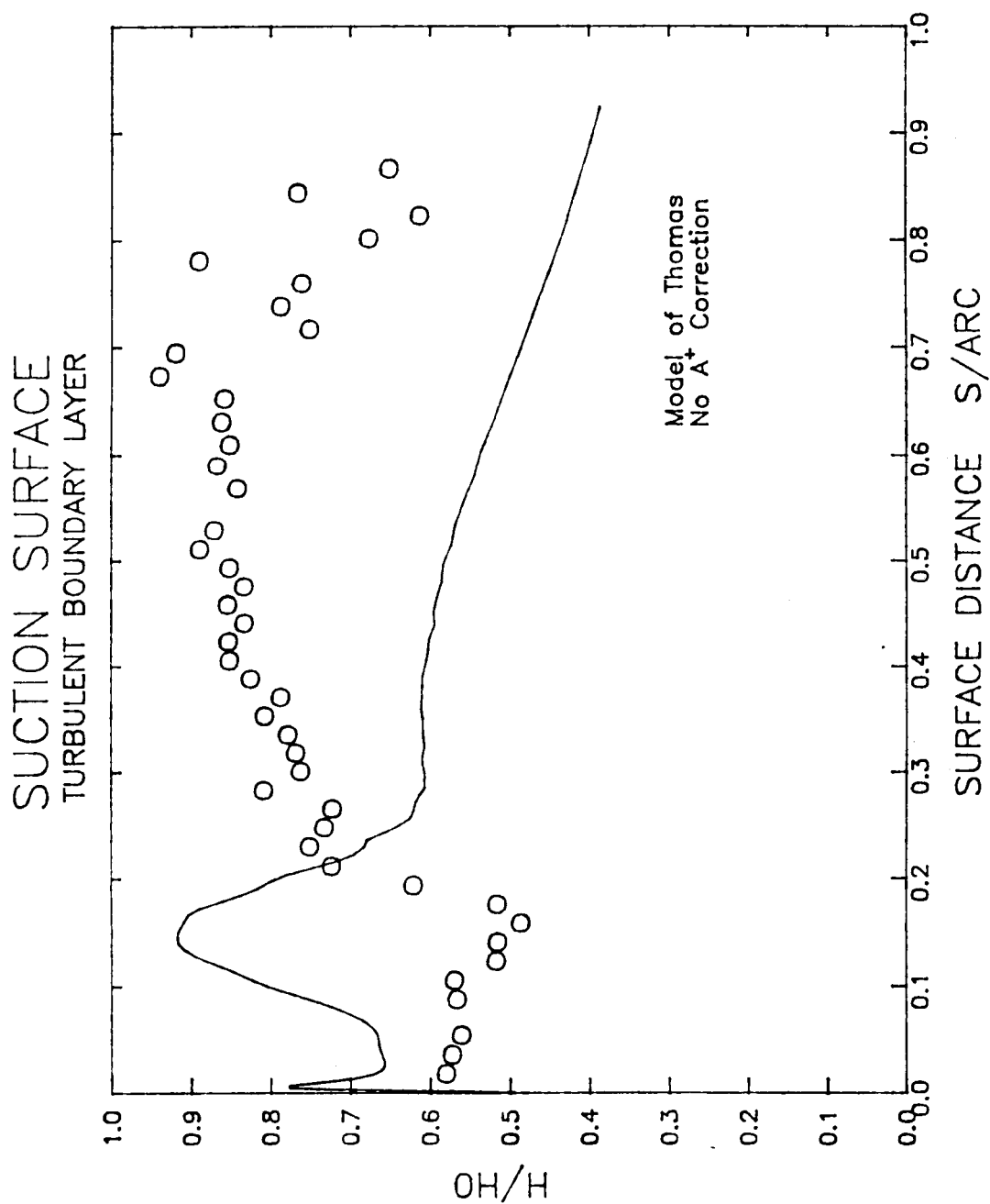


Figure 5. Heat Transfer Coefficient Distribution on the Suction Surface without using A⁺ Correction

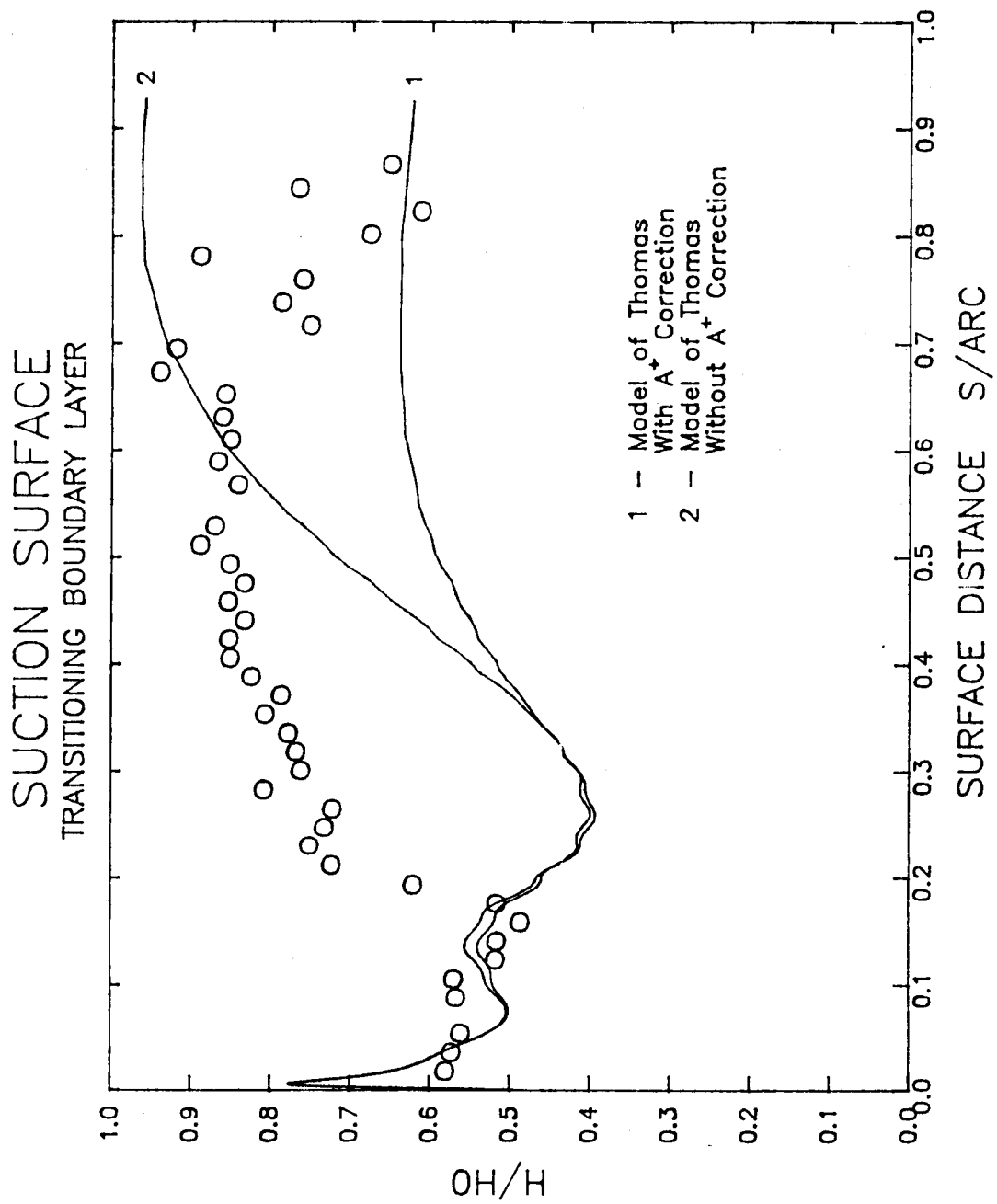


Figure 6. Heat Transfer Coefficient Distribution on the Suction Surface Assuming Transitioning Flow

CONCLUSIONS AND RECOMMENDATIONS

Based on this preliminary investigation it was concluded that the turbulent Prandtl number is not a significant force in determining the heat transfer coefficient. It appears that the character of the boundary is much more important with the turbulent Prandtl number "fine tuning" the end result. This was seen in the pressure surface results where the predictions went from poor to excellent by altering the boundary layer character via the effective viscous sublayer thickness (A^+) correction.

It is also apparent that the suction surface flow is clearly not being modeled correctly. The transition from laminar to turbulent has to date defied description via current transition models. There is certainly a complex interaction between the transition, pressure gradient, curvature, and three dimensional effects not yet understood.

Albeit preliminary, this study has suggested need for the following:

- 0 Boundary layer character on the suction surface must be understood. This suggests a detailed studied of the flow incorporating all of the important parameters such as surface curvature, flow separation, and transition.
- 0 Higher order turbulence models coupled with the turbulent Prandtl number models need to be developed. Also, the models may need to reflect any three-dimensional effects inherent in the flow over turbine blades.
- 0 The experimental data base for gas turbine engine environments must be substantially enlarged. Current prediction methods have a limited number of data by which to compare and some of the data is suspect due to the difficulty in making the measurements. Detailed measurements are necessary to validate the higher order models that appear to be needed.

REFERENCES

1. Crawford, M.E. and Kays, W.M., "STAN5 - A Program for Numerical Computation of Two-Dimensional Internal and External Boundary Layer Flow", NASA CR-2742, 1976.
2. Gaugler, R.E., "Some Modifications to, and Operational Experiences with, the Two-Dimensional, Finite-Difference, Boundary Layer Code STAN5", ASME Paper No. 81-GT-89, 1984.
3. Patankar, S.V. and Spalding, D.B., "A Finite-Difference Procedure for Solving the Equations of the Two-Dimensional Boundary Layer", International Journal of Heat and Mass Transfer, Vol. 10, 1967, p. 1389.
4. Hylton, L.D., Mihelc, M.S., Turner, E.R., Nealy, D.A., and York, R.E., "Analytical and Experimental Evaluation of the Heat Transfer Distribution Over the Surfaces of Turbine Vanes", NASA CR-168015, 1983.
5. Jenkins, R., "Variation of the Eddy Conductivity with Prandtl Modulus and its Use in the Prediction of Turbulent Heat Transfer Coefficients", Proceedings of the Heat Transfer and Fluid Mechanics Institute, 1951, p. 147.
6. Thomas, L.C., "Temperature Profiles for Liquid Metals and Moderate Prandtl Number Fluids", Journal of Heat Transfer, Vol. 92, 1970, p. 565.
7. Cebeci, T., "A Model for Eddy Conductivity and Turbulent Prandtl Number", Journal of Heat Transfer, Vol. 95C, 1973, p. 227.
8. Tyldesley, J.R. and Silver, R.S., "The Prediction of the Transport Properties of a Turbulent Fluid", International Journal of Heat and Mass Transfer, Vol. 11, 1968, p.1325.

N89-21757

1988

NASA/ASEE Summer Faculty Fellowship Program

Marshall Space Flight Center
The University of Alabama

Development of a Prototype Commonality Analysis Tool
for use in Space Programs

Prepared by:	Dorian P. Yeager
Academic Rank:	Associate Professor
University and Department:	The University of Alabama Computer Science
NASA/MSFC:	
Laboratory:	Systems Analysis and Integration
Division:	Space Station Systems
Branch:	Systems Integration
NASA Colleague:	L. Dale Thomas
Date:	August 12, 1988
Contract No.	The University of Alabama NGT-01-002-099

ABSTRACT

A software tool to aid in performing commonality analyses, called Commonality Analysis Problem Solver (CAPS), was designed, and a prototype version (CAPS 1.0) was implemented and tested. CAPS 1.0 runs in an MS-DOS or IBM PC-DOS environment. CAPS is designed around a simple input language which provides a natural syntax for the description of feasibility constraints. It provides its users with the ability to load a database representing a set of design items, describe the feasibility constraints on items in that database, and do a comprehensive cost analysis to find the most economical substitution pattern.

ACKNOWLEDGMENTS

The author wishes to thank NASA and ASEE for their support of a fine program. My experiences here have been extremely rewarding. Thanks also to Mike Freeman and Ernestine Cothran for their capable administration and their personal involvement in the program.

Special thanks go to my NASA colleague, L. Dale Thomas, without whose help the fundamental nature of the problem would not have been evident.

Thanks go also to Charlie Cothran and the crew at EL83, now EJ12, who again made me feel very much at home and provided a pleasant and productive work environment.

INTRODUCTION AND OBJECTIVES

Commonality is an attribute of large systems. A system is said to incorporate a high level of commonality if there are few instances where functionally similar but unique designs are incorporated into the system, and component parts are designed with a high level of functionality, so that each separately designed component may be employed in multiple positions within the system.

Commonality Analysis is a process, as yet only rather poorly defined, for assessing the level of commonality in the design of a large system and, if necessary, making recommendations for increasing that level of commonality. The major considerations in making such recommendations are of necessity economic ones, and it must be true in some sense that the system will be more cost-effective if the recommended changes are made.

An automatic tool for structuring the commonality analysis process will always fall short of the goal of providing a comprehensive framework. There may indeed be more art than science involved in the process. A good tool is very definitely needed, nonetheless. A first step in the direction of providing such a tool is represented by the SCAT program developed under Work Package 1 (see MSFC 1987). An assessment of the benefits and drawbacks of that program is given in last year's report, along with recommendations for improvement. Some of those recommendations are now being incorporated into SCAT.

This report describes an alternative approach to commonality analysis. The Commonality Analysis Problem Solver, or CAPS, is conceived as a software tool with two major features: (1) more intelligent algorithms for investigating substitution strategies, and (2) a language for precisely describing substitution constraints. The mathematics which suggested the CAPS algorithms was discovered by the author during his 1987 tenure as a Summer Faculty Fellow. The design and implementation of the CAPS language was the purpose of this year's work, and its description is the subject of this report.

CAPS 1.0

CAPS 1.0 is a demonstration prototype consisting of about 2800 lines of C code developed in a four-week period. Some limitations were imposed by the short development time. Notable among these limitations are the restriction to numeric data fields only, the rather simplified input data file format, and the inability to short-circuit prohibitively long cost analyses. The last-mentioned drawback derives from CAPS 1.0's insistence on providing the absolute optimum solution, no matter what the cost in computation time. Future development of CAPS will give it the intelligence to choose and apply a suboptimal solution strategy when such a strategy is indicated. Future releases will also extend CAPS' functionality and improve its user interface.

DATA FILES

A CAPS 1.0 data file is organized as a series of lines of text. The first line contains a series of field names (attributes) applicable to records in the data base. The syntax of field names must conform to the same constraints as that of CAPS variable names (see below). The remainder of the file is a series of records, coded as one line of text per record. Each record consists of a series of numeric constants, and each record must supply a value for each attribute. The form of the file is therefore exactly like a table with headings, except no alignment conventions need be observed. The data file is a simple ASCII text file, such as could be constructed with any text editor, including the DOS EDLIN editor, or with any word processor capable of exporting documents to an ASCII file. Alternatively, a CAPS 1.0 file can easily be generated as a report file by any relational database. An example of a data file is given below:

ddt&e	prod	weight	volume	quantity
46.166	7.694	9.222	19.2	1
49.374	8.229	10.375	21.6	1
67.833	11.306	17.292	36.0	4
71.86	11.977	19.021	39.6	4
92.819	15.47	28.244	58.8	2
355.772	59.295	232.289	483.6	2
366.685	61.114	243.241	506.4	6
378.240	63.040	253.616	528.0	3
464.314	77.386	348.722	726.0	4

Table 1: Typical CAPS data file.

In the above example, there are five fields and nine records. Note that no quote marks are used to delimit the field names. In CAPS 1.0, no input line can be longer than 300 characters.

THE CAPS 1.0 USER INTERFACE

CAPS 1.0 uses a command line input scheme. When the program initially loads, it displays the prompt:

>>

CAPS then recognizes commands typed in by the user and responds to each command in turn by displaying (a) the value of an expression, (b) a graph of the data set, or (c) a report of current status or of an action taken. There are 16 input forms, as follows:

```
<expression>
load <file_name>
fields
for <logical expression> allow <logical expression>
for <logical expression> disallow <logical expression>
allow all
disallow all
status
define <variable> as <expression>
add <defined variable>
graph <field_name_1> vs <field_name_2>
cost
learn <constant>
learn
take <file_name>
quit
```

We will discuss each of these in turn below.

The following keywords are reserved words in CAPS 1.0. Their meanings are fixed and cannot be changed by the user, which means that they are not available as field names in data files nor as variable names.

abs	add	all	allow	and	as	cos	cost
define	disallow	exp	fields	for	graph	int	item
learn	ln	load	log	not	or	quit	sin
sqr	status	take	tan	vs			

CAPS EXPRESSIONS

The primary use for CAPS expressions is the description of feasibility constraints for the substitution strategy in commonality analyses. In order to describe such constraints it must be possible to communicate any kind of relationship between two items in a data file. Expression syntax is essential for such communication.

CAPS expressions use standard expression syntax, like that used in FORTRAN and BASIC. Besides the standard arithmetic, relational, and logical operators, several essential functions are implemented, such as sine, cosine, tangent, square root, absolute value, and integer part.

Spaces are ignored wherever they occur in CAPS expressions, except that spaces are not allowed in variable names or numeric constants.

NUMERIC CONSTANTS

CAPS numeric constants may be integers or real numbers in fixed or floating point form, and may be signed or unsigned. Only decimal constants are allowed. Examples are as follows:

25	-340	.7	0	0.789	8.98E23
345.	-.001	1e2	1e-2	7.38E-21	500000

Very few restrictions of commonly accepted syntax are applicable. The range of values accepted is $-3.4E38$ to $3.4E38$. Numbers smaller in magnitude than $1.2e-38$ underflow to zero. Precision is about seven decimal digits. CAPS does not accept floating point constants with more than two digits in the exponent. Thus $1e010$ is not a valid constant, but $1e01$ and $1e10$ are valid.

In CAPS, all numeric quantities are represented internally as real (floating point) numbers.

VARIABLE NAMES

CAPS variable names may be of any length, must begin with an alphabetic character, and must consist of alphabetic characters, decimal digits, underline characters (), and ampersands (&). CAPS is NOT case-sensitive. Thus the names VOL, Vol, vol, and VO1 all refer to the same variable.

ARITHMETIC OPERATORS

CAPS uses the FORTRAN set of arithmetic operators, with the same meaning and the same precedence rules, given below:

Operator	Precedence	Meaning
+	Lowest	Real number addition
-	Lowest	Real number subtraction
*	Intermediate	Real number multiplication
/	Intermediate	Real number division
**	Highest	Exponentiation

Since all numbers are represented internally in floating point, an operation performed on two integers will not necessarily yield an integer. For example, the expression $1/3$ evaluates to 0.3333334, not 0 as it would in FORTRAN.

Exponentiation is allowed whenever it makes sense and the result is a real number. Thus $0**0$ is not permitted, nor is $(-1)**0.5$, but $(-2)**(-3)$ is permitted and yields -0.125.

The order of evaluation of operators with the same precedence depends on their associativity. The associativity of +, -, *, and / is from left to right, whereas that of ** is from right to left. This is exactly the convention observed by FORTRAN. Associativity and precedence rules may be overridden with the use of parentheses. Some examples follow:

Expression	Value
$25 - 12 + 2$	15
$25 - (12 + 2)$	11
$5 / 2 * 2$	5
$5 / (2 * 2)$	1.25
$2 ** 2 ** 3$	256
$(2 ** 2) ** 3$	64

Negation, or unary minus, is a separate operation, different from subtraction. Its precedence is higher than that of the additive operators + and - and lower than that of the multiplicative operators * and /. Thus the expression $4 + -5$ is allowed, but the expression $2 ** -2$ is not valid CAPS syntax. Parentheses are required around a negated CAPS expression which is used as a right-hand input to any operation other than + or -. Thus in the example it is necessary to use the syntax $2 ** (-2)$.

PRIMITIVE FUNCTIONS

CAPS 1.0 supports the following set of primitive functions:

Function Name	Meaning
sqrt	Square root
abs	Absolute value
sin	Trigonometric sine
cos	Trigonometric cosine
tan	Trigonometric tangent
ln	Logarithm base e
exp	Exponential function (e^x)
log	Logarithm base 10
int	Integer part

Most of the above have the obvious meaning, but a few explanations are needed. (1) The three trigonometric functions take their operands in radians, not degrees. (2) CAPS intercepts any attempts to use one of these functions with an inappropriate argument. For example, CAPS will print its own error message in response to `sqrt(-2)`, and will display again the ">>" input prompt. In contrast, some other types of floating point errors, such as floating point overflow, will not be intercepted by CAPS and will cause immediate program termination. (3) The 'int' function returns the integer part of a number. Examples follow:

Expression	Value
<code>int(5.7)</code>	5
<code>int(-5.7)</code>	-5
<code>int(3)</code>	3

RELATIONAL OPERATORS

The CAPS relational operators are `=`, `<>`, `<`, `>`, `<=`, and `>=`. They are interpreted, respectively, as 'is equal to', 'is not equal to', 'is less than', 'is greater than', 'is less than or equal to', and 'is greater than or equal to'. The relational operators all have lower precedence than the arithmetic operators. Since the output of a relational expression is not normally given in turn as input to another relational operation (for example, `5<7<9` is not a valid CAPS expression), associativity rules are not needed.

A CAPS relational expression always returns a value of 1 or 0, meaning true or false respectively. Thus the expressions

5<7 and -7.3<-5.9 return the value 1 and the expressions 5>=7, 8=9, 1e2<99, and 3<>3 all return the value 0.

Parenthesized expressions are fair game for input into any CAPS relational operation. Thus (5<7)<9 is legal and yields in turn first 1<9, then 1, i.e. true.

BOOLEAN OPERATORS

The boolean operators are given, along with their precedences, in the table below:

Operator	Precedence
or	lowest
and	intermediate
not	highest

The precedence rules were chosen in order to observe standard conventions with almost universal acceptance. All boolean operators have lower precedence than all relational operators and all arithmetic operators. Thus

$x < 3$ and not $y \geq 4$ or $p = q$

is equivalent to

$((x < 3) \text{ and } (\text{not } (y \geq 4))) \text{ or } (p = q)$

All numeric results have a true-false interpretation. Very simply, zero is false and all nonzero numbers are true. Thus '8 or 0' has the value 1, or true, whereas '8 and 0' has the value 0, or false.

EVALUATION ORDER

Evaluation order of a CAPS expression is very precisely specified, as follows: (1) with the exception of unbound variables used as operands of '=', all operands of an operator are evaluated to yield a numeric value before the operation itself is executed; and (2) the left-hand operand is always evaluated before the right-hand operand.

VARIABLE BINDINGS AND THE ASSIGNMENT SIDE EFFECT

A CAPS variable, just like a variable in any standard programming language, may be associated with a value. Until

such an association takes effect, the variable is said to be unbound. A variable may be bound to a value using the = operator.

The = operator, being a relational operator, always returns as its value the number 1 or 0, indicating a true or false result. If one of the operands of = is an unbound variable, two things occur. First, the value of the other operand is computed and bound to the variable. Second, the value 1 is returned, since as a result of the binding that has taken place the two sides of the = operator have identical values.

Consider, for example, the expression below, in which the variable x is initially unbound:

4 = x and x > 3

The left-hand operand of the 'and' operation (i.e. 4 = x) is first evaluated, having the dual effect of assigning 4 to x and returning a true value. The right hand operand is then evaluated as 4 > 3, also yielding a true value. Since both operands are true, the 'and' operation also yields a true value, and along the way x picks up its binding to the value 4.

CALCULATOR OPERATION

Besides its primary use for commonality analyses, CAPS may be used as an interactive calculator. This feature was included in order to allow the user to experiment with CAPS expression evaluation to better understand its semantics. Calculator operation is achieved when the user types in an expression and CAPS evaluates the expression and displays its value. Consider, for example, the following series of CAPS inputs and responses:

```
>>principle = 40000
1.000000
>>rate = 0.10
1.000000
>>rate/12 = i
1.000000
>>principle*i / (1 - (1+i)**(-360)) = payment
1.000000
>>payment
351.028687
>>
```

Notice that the first four inputs are assignments, which always return the value 1. The last input is a single variable name,

'payment'. CAPS checks the symbol table for its current value, and displays that value as its response.

THE SYMBOL TABLE

The central data structure internal to CAPS, used for maintaining current variable bindings, is its symbol table. During a CAPS interactive session, all symbols which have any meaning to CAPS are represented in the symbol table. This includes the keywords listed above as well as all variables which have been used in previous commands and all field names from the currently loaded data set.

When CAPS is initially invoked, the only symbols represented in its symbol table are the keywords. As variables are introduced they are stored in the symbol table along with their bindings, if any. Once a variable is given a binding, that binding cannot be changed until a new data file is loaded. When a new file is loaded, the symbol table is cleared of all symbols except for the keywords and the field names for that file.

THE 'LOAD' COMMAND

The syntax of the load command is

```
load <file name>
```

The rules of formation for CAPS 1.0 file names are exactly like those for CAPS variables. The DOS file name extension for CAPS 1.0 data files is always 'dat', and is provided automatically. The user is not allowed to provide an alternate extension. Since DOS requires that file names be limited to eight characters, CAPS automatically truncates any excess trailing characters from names which are longer than this limit.

CAPS' actions upon receiving a load command are to (a) clear the symbol table of all variable bindings, (b) read the field names from the file and install them in the symbol table so that they may not be used for variable names, and (c) read the data from the file into an internal table.

If during the loading of a file an error is discovered by CAPS in the way the file is organized, CAPS displays an error message and aborts the load process. A side effect of any load is the clearing of the field names of the last loaded file from

the symbol table. This means that after an unsuccessful load attempt CAPS will have no memory of the last file successfully loaded.

When a file is successfully loaded, CAPS reports the number of fields and the number of items of data. Following is an example.

```
>>load motors
12 fields, 34 items.
Load successfully completed.
>>
```

LIMITATIONS: CAPS 1.0 accommodates a maximum of 75 items and 256 fields.

THE 'FIELDS' COMMAND

The fields command requests a listing of all the field names for the currently loaded data set. For example:

```
>>fields
ddt&e unitcost volume
3 fields.
>>
```

THE FEASIBILITY MATRIX

There are three internal data structures which CAPS uses to reflect its current state. The first is the symbol table, and the second is the table containing the currently loaded data set. The third is the feasibility matrix, called alpha, which determines which substitutions are currently permitted. Alpha is boolean and is doubly indexed over the items in the data set. Thus α_{ij} is true if item i is allowed to substitute for item j, false otherwise.

COMMANDS USED TO ALLOW AND DISALLOW SUBSTITUTIONS

After a file has been loaded containing information on the items to be subjected to commonality analysis, it is necessary to communicate to CAPS the allowable patterns of substitution. This communication is achieved as follows: Initially, CAPS assumes no substitutions are allowed. This means that the feasibility matrix has 'true' entries only on the diagonal. Then, using a series of 'for', 'allow all', and 'disallow all' commands, the user modifies the feasibility matrix. When the

user has adequately described all allowable substitutions, s/he requests a cost analysis with the 'cost' command.

THE 'FOR' COMMANDS

The for commands are:

for <expression> allow <expression>

and

for <expression> disallow <expression>

The first allows new substitutions. It does not, of course, require that those substitutions be made. It simply tells the cost analysis portion of the program that it is permitted to consider the indicated substitutions as options when it performs its analysis. The second form forbids the cost analyzer to consider a given set of substitutions.

Examples:

```
for capacity=10 allow capacity=12
for boys_type and speeds=3 allow boys_type and (speeds=3 or
speeds=10)
for not flammable disallow flammable
for portable disallow not portable
```

In the above, 'capacity', 'boys_type', 'speeds', 'flammable', and 'portable' are field names. The first statement allows all items for which the value in the field 'capacity' is 12 to substitute for all items having the value 10 in that field. The second can be summarized rather clearly if we think in terms of a data base representing bicycles. If for a given bicycle in the data base the boolean field 'boys_type' is set to true and the value of 'speeds' is 3, then any bike used as a substitute for that bike must also be a boys bike and must be either a 3-speed or a 10-speed.

The last two examples also involve boolean fields, but are phrased in the negative. Example 3 says that a flammable item may not substitute for an item which is not flammable, and example 4 says that an item which is not portable may not substitute for an item which is portable.

Note that disallowing substitutions has no effect until some substitutions have first been allowed, since CAPS initially assumes no substitutions are valid.

THE 'ALLOW ALL' AND 'DISALLOW ALL' COMMANDS

The command 'allow all' tells CAPS all substitutions are valid, whereas 'disallow all' clears the feasibility matrix (except, of course, for the diagonal) and puts CAPS back in its original state with respect to the currently loaded table.

USE OF VARIABLES

To provide as general a facility as possible for the description of substitution constraints, two alternative kinds of variable bindings in for statements (other than simple bindings to numerical values) may be employed. Firstly, a variable may be bound to a field name in the currently loaded table, and secondly it may be bound to an expression involving one or more field names. The example below will illustrate:

```
for size = x allow size >= x
```

This example basically says that any item of a given size may be replaced by an item of that size or larger. The actual procedure CAPS goes through is as follows.

Clear from the symbol table all variables having numeric bindings.

For each item i in the table:

 Bind x to the size of item i

 For each item j in the table:

 Allow j to substitute for i provided the size of j
 is

 at least x

 Free x's binding

Consider now a second example:

```
for x=power/weight and x>=1.5 allow y=power/weight and y>=x
```

Here two variables named x and y are employed. The for statement says that if x is 'power' divided by 'weight' for item i, x is at least 1.5, and y is 'power' divided by 'weight' for item j, then j can replace i provided y is at least as large as x. The following is an alternate formulation of the same set of allowed substitutions, using only a single variable:

```
for x=power/weight and x>=1.5 allow power/weight>=x
```

THE KEYWORD 'ALL'

When used in 'for' statements, the keyword 'all' refers to all items in the currently loaded data set. For example:

```
for cylinders=4 allow all
```

allows all items in the data set to substitute for any item having the value 4 in its 'cylinders' field.

In particular, 'for all allow all' has the same effect as 'allow all', and 'for all disallow all' has the same effect as 'disallow all'.

THE 'DEFINE' COMMAND

The 'define' command has the form:

```
define <variable> as <expression>
```

This statement binds the variable to an expression, not to a value. It allows the user to establish meanings for variables which go beyond simple numeric bindings and which allow more brevity and flexibility in the coding of meaningful for statements. Also, since the for statement destroys numeric bindings, the define statement provides a way of protecting bindings from the effects of a 'for'.

Examples are as follows:

```
define pi as 3.1415926
define totalcost as
  weight*cost_per_pound+volume*cost_per_cubic_foot
```

Note in particular the first example, where pi is bound to an expression consisting only of the constant 3.1415926. Although the expression will always yield that constant as its value, it is nevertheless stored internally as an expression and is therefore protected from the effects of for statements. A defined variable may be used in subsequent expressions in for statements. If the variable being defined is not defined in terms of field names, it may be used in calculator mode. In that case, a numerical result is computed for the variable using the current bindings of other variables in the expression. The following CAPS dialog will illustrate.

```
>>define pi as 3.141593
Definition successful.
>>define angle1 as pi/6
Definition successful.
```

```
>>sin(angle1)
0.500000
```

Since 'define' does not immediately evaluate the expression provided as its second operand, but waits until the defined variable is used in calculator mode or in a for statement, the following is also possible.

```
>>define angle1 as pi/6
Definition successful.
>>define pi as 3.141593
Definition successful.
>>sin(angle1)
0.500000
```

The for statement above can be rephrased using a define statement, as follows:

```
define rel_power as power/weight
for x=rel_power and x>=1.5 allow rel_power>=x
```

THE KEYWORD 'ITEM'

When used in 'define' and 'for' statements, the keyword 'item' refers to an internal counter used to number the items in the table. Items in the table are automatically numbered from 1 to n, where n is the total number of items in the table. The status of 'item' is similar to that of a field name, but its value is not explicitly coded into the table. Examples:

```
for item=7 allow all
for item=x allow item>x
define even as int(item/2)*2 = item
```

The first example allows all items to replace the seventh item in the table. The second allows any item to be replaced by any item following it in the table. The third associates the variable 'even' with an expression which evaluates to true if and only if the item is numbered with an even number.

THE 'STATUS' COMMAND

The status command reports on the current status of the feasibility matrix by indicating which items each element of the data set is currently allowed to substitute for.

Example:

```
>>load data1
4 fields, 5 items.
Load successfully completed.
>>status
Allowable substitutions:
  1 -> {1}
  2 -> {2}
  3 -> {3}
  4 -> {4}
  5 -> {5}
>>for item=x allow item>x
10 new substitutions allowed.
>>status
Allowable substitutions:
  1 -> {1}
  2 -> {1,2}
  3 -> {1,2,3}
  4 -> {1,2,3,4}
  5 -> {1,2,3,4,5}
>>
```

HISTORY SENSITIVITY

The user of CAPS should always be aware that the purpose of a single for command is to augment or restrict the set of allowable substitutions, not to make an isolated declaration. CAPS is designed to communicate these allowed substitutions via a series of history-sensitive operations, the results of which are recorded in the feasibility matrix alpha.

Consider the following set of commands, involving valves:

```
define rel_diff as abs(diameter-x)/x
for diameter=x allow rel_diff < 0.05
for gas disallow liquid
```

The effect of the above is radically different from the following:

```
for gas disallow liquid
define rel_diff as abs(diameter-x)/x
for diameter=x allow rel_diff < 0.05
```

Note that the only difference in the two sets of commands is the order in which the commands were given. But because of the history-sensitive nature of CAPS, the results are quite different. In the first example no liquid valve will be allowed to substitute for a gas valve. But in the second

example two valves whose diameters differ by 5% or less will be allowed to substitute one for the other, regardless of whether they are liquid or gas valves.

THE 'ADD' COMMAND

The add command is used for adding new fields to a loaded data set. A variable may be used to add such a field only if it was bound using a 'define' statement. An example follows:

```
add rel_power
```

Here the variable rel_power is added as a new field. The mechanics of this process are simple - the size of the internal table used to hold the data is increased by one column, and data is placed in that column by stepping through the table and computing the value of the expression associated with rel_power for each data item in the table.

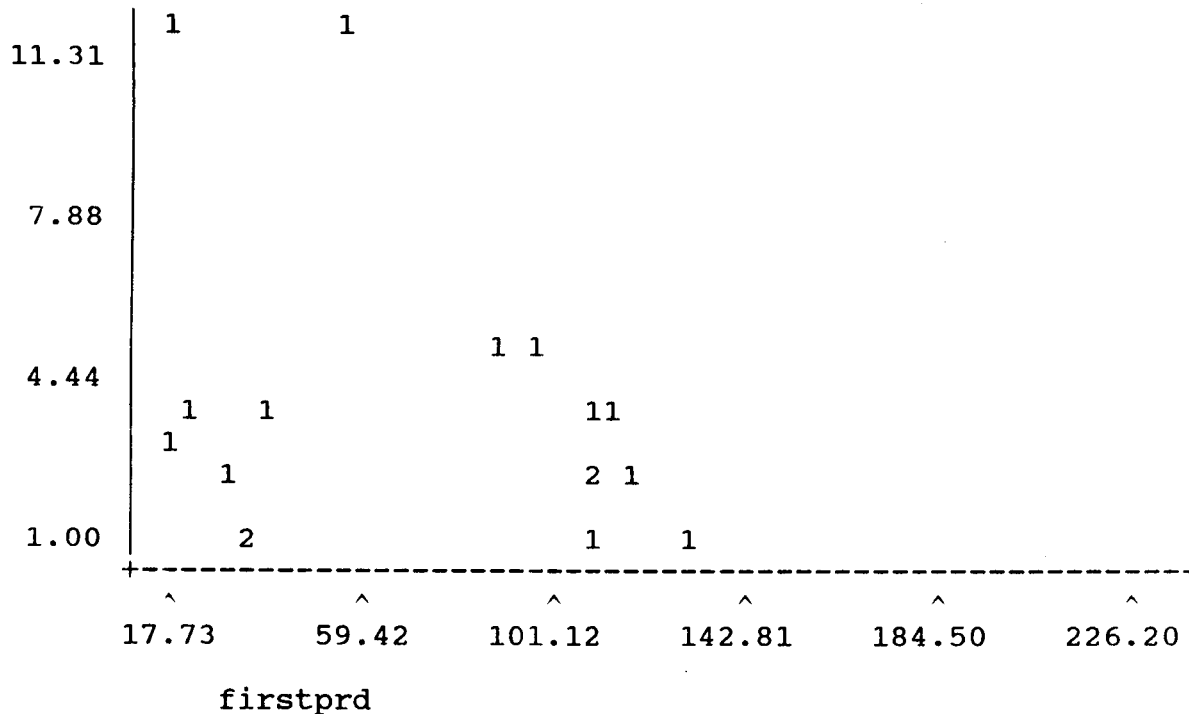
Note that the variable is now no longer associated with the expression. It has become a field name. In database terms it is a derived attribute based on the values of other attributes, or fields.

There are three reasons for converting defined variables into field names with the 'add' command. The first is increased speed of execution of 'for' commands. If a defined variable is used in the second operand of a for statement on a table with 30 elements, then the expression associated with that variable will be evaluated 900 times. Expression evaluation is much more expensive than table lookup, and installing the variable as a derived attribute requires only 30 evaluations of the expression. The second reason for adding new fields is to be able to use those fields in graphs (see below). The third reason is to provide necessary parameters for a cost analysis.

THE 'GRAPH' COMMAND

The graph command allows the user to take graphic snapshots of the data. CAPS 1.0 uses the text screen to provide a low-resolution graph of the data set. Two field names are supplied as parameters for the purpose of labelling the horizontal and vertical axes. An example follows.

```
>>graph firstprd vs quantity
quantity
```



```
>>
```

In the example, there are two items having quantity = 1 and having a value for firstprd between 17.73 and 59.42. There are five with quantity between 2 and 4 and firstprd between 101.12 and 142.81. CAPS 1.0 displays a digit between 1 and 9 in a given position if there are fewer than ten items falling into that position of the graph. If there are more than 9 data points, CAPS displays an asterisk (*) in that position.

THE 'COST' COMMAND

The cost command requests a cost analysis. In order for a meaningful cost analysis to be conducted, CAPS must have been supplied with a loaded data set containing one of the following sets of attributes:

1. ddt&e, quantity, unitcost, and firstprd.
2. ddt&e, quantity, and firstprd.
3. ddt&e, quantity, and unitcost.

CAPS must also have been supplied with a valid feasibility matrix via a series of 'for', 'allow all', and 'disallow all' commands.

The ddt&e field is the design, development, test, and engineering costs associated with the item. It is an "up front" cost, paid only once no matter how many copies of the item are to be produced. The ddt&e field is essential for any kind of cost analysis in CAPS 1.0.

The quantity field is the total number of copies of the item which will be produced. For manufacturing applications this tends to be quite large, but for space programs it is usually a relatively small number. Like the ddt&e field, the quantity field is essential for CAPS 1.0 cost analyses.

The unitcost field is the sum of all costs associated with each new copy of an item which are not subject to a learning curve. Its value is multiplied by the value in the quantity field. The firstprd field is the first product cost of an item. If this field is present in the data set, CAPS 1.0 will use the learning curve cost function.

The two cost functions employed by CAPS 1.0 are:

1. The linear cost function is used if firstprd is not present. With this model the cost associated with a given item is

$$\text{ddt\&e} + \text{quantity} * \text{unitcost}$$

2. The learning curve cost function is used if firstprd is present. Here the cost associated with a given item is

$$\text{ddt\&e} + \text{quantity} * \text{unitcost} + \text{firstprd} * \sum_{i=1}^n i^t,$$

where $n = \text{quantity}$ and $t = \ln(L/100)/\ln(2)$. Here L is the learning curve parameter, adjustable in CAPS via the 'learn' command.

THE 'LEARN' COMMAND

It is possible to communicate the value of the learning curve parameter to CAPS' cost analysis component. This is a percentage, usually around 85%, but the value varies with the type of item being subjected to analysis. For example,

learn 90

sets the learning curve at 90%. CAPS will assume a learning curve of 85% until told to change that value.

If the learn command is used without an argument, CAPS displays the current value of the learning curve parameter.

THE 'TAKE' COMMAND

The 'take' command allows the CAPS user to prepare a series of commands and store them in an ASCII file. This avoids having to reenter the same or nearly the same series of commands each time s/he subjects the same data set to a cost analysis. For example, in response to the command

```
>>take bicycles
```

CAPS reads the commands in the file 'bicycles.tak' and executes them one by one. The 'tak' extension is always assumed.

INTERPRETING THE COST ANALYSIS OUTPUT

Following is a typical CAPS 1.0 cost analysis:

```
>>cost.
```

Time required for an exhaustive analysis: approximately 280 seconds.

CAPS 1.0 will find the optimum solution in approximately 3 seconds.

Proceed? y

Working.....

Item	Replaces	Replacement Cost	Unique Cost
3	{1,2,3}		
		326.77	401.09
4	{4}		
		228.97	228.97
7	{5,6,7,8}		
		3389.18	4299.85

TOTALS		3944.92	4929.91

```
>>
```

The analysis above is for a data set of eight items. CAPS begins by giving an estimate of the time which would be required for an exhaustive analysis, followed by an estimate of the time for it to do its own analysis. The second number is, of course, never greater than the first. It is, however, quite often much smaller because CAPS employs its knowledge about the linear cost function and the learning curve cost function to attempt to reduce the number of potential solutions to be examined. (See Yeager, 1987a and 1988) The estimates are very rough, and are based on IBM PC/AT run times and the complexity of the algorithms being used.

In the example, CAPS recommends that only three of the items, items 3, 4, and 7, be produced. Items 1 and 2 are to be replaced by item 3, and items 5, 6, and 8 by item 7. Item 4 is a unique item. In making its recommendation, CAPS has followed the feasibility constraints placed on the data by the user in previous commands.

CAPS 1.0 displays the "Working...." message to give the user an indication of the progress of the analysis. Each dot ('.') represents three potential solutions (partitions) examined.

CAPS 1.0 cost figures are for comparison only. CAPS is intended as a tool for Commonality Analysis, not for projection of costs. It is, of course, true that the better the cost estimates provided to the 'ddt&e', 'unitcost', and 'firstprd' fields, the more accurate will be the recommendations of CAPS.

THE 'QUIT' COMMAND

The quit command returns control to DOS. Since no provision is made in the CAPS 1.0 prototype for saving results of an analysis to a file, the user is advised to use 'take' files to collect all relevant commands pertaining to a given data base, and to direct output to the printer during important analyses.

CASE STUDY #1

The first case study is for a set of storage tanks, the data for which appears in Table 1. The tanks appear in order by size, and size is the only determinant for substitutability. If the data is stored in file "tanks1.dat", the following CAPS dialogue yields a linear cost analysis.

```
>>load tanks1
5 fields, 9 items.
Load successfully completed.
>>fields
ddt&e prod weight volume quantity
5 fields.
>>define unitcost as prod+weight+volume
Definition successful.
>>add unitcost
Field successfully added.
>>for item=x allow item>x
36 new substitutions allowed.
>>cost
```

Time required for an exhaustive analysis: approximately 630 seconds.

CAPS 1.0 will find the optimum solution in approximately 1 seconds.

Proceed? y

Working.....

Item	Replaces	Replacement Cost	Unique Cost
2	{1,2}	129.78	171.86
4	{3,4}	636.64	680.48
5	{5}	297.85	297.85
8	{6,7,8}	9669.46	10049.56
9	{9}	5072.75	5072.75
<hr/>			
TOTALS		15806.47	16272.49

CASE STUDY #2

The second case study also deals with storage tanks, but here a distinction is made between 'liquid' and 'gas' tanks. Specifically, we wish to enforce that only liquid tanks replace liquid tanks and that only gas tanks replace gas tanks. The data is in table 2.

tank	ddt&e	unitcost	firstprcd	quantity	size	liquid	gas
1	148.65	167.52	29.73	2	2	1	0
2	549.65	1302.46	109.93	2	5	1	0
3	549.65	1302.46	109.93	4	5	1	0
4	88.64	74.43	17.73	3	0	1	0
5	588.91	1451.32	117.78	2	6	1	0
6	102.20	93.08	20.44	4	1	1	0
7	178.57	223.29	35.71	1	3	1	0
8	178.57	223.29	35.71	1	3	1	0
9	663.29	1750.64	132.66	1	7	1	0
10	549.65	1302.46	109.93	2	5	1	0
11	549.65	1302.46	109.93	1	5	1	0
12	579.24	3787.29	115.85	4	8	1	0
13	192.50	671.48	38.50	4	1	0	1
14	286.79	1257.69	57.36	12	4	1	0
15	95.29	223.83	19.06	12	0	0	1
16	1318.61	13767.10	263.72	5	9	1	0
17	438.09	2451.43	87.62	5	2	0	1

Table 2. Liquid and gas storage tanks input file.

>>load tanks2

3 fields, 17 items.
 Load successfully completed.
 >>for liquid and size=x allow liquid and size>=x
 98 new substitutions allowed.
 >>for gas and size=x allow gas and size>=x
 3 new substitutions allowed.
 >>cost
 Time required for an exhaustive analysis: approximately
 605404800 seconds.
 CAPS 1.0 will find the optimum solution in approximately 12
 seconds.
 Proceed? y
 Working.....

Item	Replaces	Replacement Cost	Unique Cost
7	{1,7,8}	1191.19	1413.83
2	{2,3,10,11}	12989.99	14805.17
5	{5}	3709.44	3709.44
9	{9}	2546.59	2546.59
6	{4,6}	862.54	901.33
12	{12}	16115.97	16115.97
13	{13}	3007.22	3007.22
14	{14}	15851.97	15851.97
15	{15}	2938.39	2938.39
16	{16}	71217.19	71217.19
17	{17}	13048.44	13048.44

TOTALS		143478.93	145555.55

CASE STUDY #3

The final example relates to interface plates for the
 modular racks used to organize work, storage, and living space
 in space station modules. Each plate can accommodate zero or
 more utility interfaces, chosen from the following set:

avionics
 nominal_power
 high_power


```
fire_detection
data_management
thermal_control
hygiene_water
nitrogen
potable_water
hygiene_waste
```

Each of the above is represented in the data set by a boolean attribute of the same name. For a given interface plate, the utility interface is present on that plate if its corresponding attribute has the value 1, and not present if that attribute has the value 0. The major constraint in this example is that no interface plate be allowed to substitute for another if the latter contains an interface which the former does not have. The following "take file" was prepared to define these constraints for CAPS:

```
allow all
for avionics disallow not avionics
for nominal_power disallow not nominal_power
for high_power disallow not high_power
for fire_detection disallow not fire_detection
for data_management disallow not data_management
for thermal_control disallow not thermal_control
for hygiene_water disallow not hygiene_water
for nitrogen disallow not nitrogen
for potable_water disallow not potable_water
for hygiene_waste disallow not hygiene_waste
```

A relatively simple approach to cost is to count one monetary unit for each interface included on a given plate. With this approach we can define the 'unitcost' field by adding the following lines to the 'take file':

```
define unitcost as avionics+nominal_power+high_power+
  fire_detection+ data_management+thermal_control+
  hygiene_water+nitrogen+ potable_water+hygiene_waste
add unitcost
```

The data used for the example is not reproduced here. It was simply an arbitrary matrix of 1's and 0's with the 'ddt&e' field set to 11 monetary units for each item. The following analysis results:

>>status

Allowable substitutions:

1 -> {1,6,7,8}
2 -> {2,3,8}
3 -> {3}
4 -> {4}
5 -> {4,5,9,10,11}
6 -> {6}
7 -> {7,8}
8 -> {8}
9 -> {9}
10 -> {9,10}
11 -> {11}

>>cost

Time required for an exhaustive analysis: approximately 17 seconds.

CAPS 1.0 will find the optimum solution in approximately 17 seconds.

Proceed? y

Working.....

Item	Replaces	Replacement Cost	Unique Cost
1	{1,6,7}	38.00	54.00
2	{2,3,8}	26.00	43.00
5	{4,5,9,10,11}	51.00	80.00

TOTALS		115.00	177.00

The first two examples are nearly identical to the two case studies used in Yeager, 1987a and 1988. The third is a deliberately scaled down version of a study in Thomas, 1988. The scaling down was necessary because of CAPS' exhaustive solution strategy. This strategy will be compromised in future versions of CAPS, so that CAPS will be able to produce a suboptimal solution to any size problem and optimal solutions to those which have good properties, such as those given here.

CONCLUSIONS AND RECOMMENDATIONS

The CAPS 1.0 prototype was successfully designed, implemented, and tested, and conclusively demonstrates the feasibility of an intelligent software tool which can not only provide a more sophisticated approach to the solution of commonality analysis problems, but can also aid in structuring the analysis procedure itself. Future development of CAPS will increase its flexibility and functionality dramatically.

REFERENCES

- MSFC. 1987. Commonality Analysis Study, User Manual for the System Commonality Analysis Tool (SCAT), D483-10064, March 1987. Contract NAS8-36413, NASA George C. Marshall Space Flight Center, Alabama.
- THOMAS, L. D. 1988. Commonality Analysis Using Clustering Methods, submitted to IEEE Transactions on Systems, Man, and Cybernetics.
- YEAGER, D. P. 1987a. Expert System Development for Commonality Analysis in Space Programs, Final Report, NASA/ASEE Summer Faculty Fellowship Program, pp. XXXV-i ff. Contract NGT-01-008-021, NASA George C. Marshall Space Flight Center, Alabama.
- YEAGER, D. P. 1987b. Commonality Analysis as a Knowledge Acquisition Problem, Proceedings of the Third Annual Conference on Artificial Intelligence for Space Applications, November 3, 1987.
- YEAGER, D. P. 1988. A Formalization of the Commonality Analysis Problem and Some Partial Solutions, submitted to Mathematics of Operations Research.

N89-21758

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER

THE UNIVERSITY OF ALABAMA

TECHNIQUES FOR STUDYING THE EFFECTS OF MICROGRAVITY
ON MODEL PARTICLE/CELL SYSTEMS

Prepared by:	Ronald B. Young
Academic Rank:	Professor
University and Department:	Univ. of Alabama in Huntsville Department of Biological Sciences
NASA/MSFC	
Laboratory:	Space Science
Division:	Low Gravity Science
Branch:	Biophysics
MSFC Colleague:	Robert Snyder
Date:	August 19, 1988
Contract No:	NGT-01-002-099 University of Alabama

TECHNIQUES FOR STUDYING THE EFFECTS OF MICROGRAVITY
ON MODEL PARTICLE/CELL SYSTEMS

by

Ronald B. Young
Professor of Biological Sciences
University of Alabama in Huntsville
Huntsville, Alabama 35899

ABSTRACT

To study the direct effects of a low gravity environment on skeletal and cardiac muscle cells, experiments were initiated to determine whether skeletal and/or cardiac muscle cells would grow within the lumen of XM-80 hollow fibers (i.d. = 0.5 mm). Cells were prepared from skeletal or cardiac muscle tissue of 12 day embryos and were cultured for up to 7 days in the hollow fiber environment. Light microscopy revealed that cells proliferated to confluency over this period of time and fusion was apparent in the skeletal muscle cells. Once it had been verified that cells would grow to confluency, additional XM-80 fibers containing cells were placed in a Clinostat in the horizontal position at 100 rpm. Fibers were stretched by a built-in spring mechanism to hold the fiber tightly at the center of rotation. Under these conditions, the gravity vector approaches zero and the cells are in an environment that simulates microgravity. Examination of skeletal muscle cells by electron microscopy revealed that myoblast fusion and myofibril accumulation were extensive. Although data obtained thus far are preliminary, they suggest that myofibril organization in chicken skeletal muscle cultures is somewhat more poorly defined in Clinostat-rotated cultures than in controls that were not subjected to Clinostat conditions.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Robert S. Snyder for providing the opportunity for the Summer Faculty Fellowship and Ms. Teresa Miller for assistance with all aspects of the research. This work could not have been possible without the generous contributions of their time and space.

I also thank my coworkers in the Department of Biological Sciences at the University of Alabama in Huntsville for their hard work in my absence, and for contributing to the development of this project. Special thanks go to Donna Donnelly and Debbie Windham for their hard work and skill with the Clinostat and the microscopy, to Donna Brown for preparing the bovine muscle cell cultures, and to my Co-Investigator, Dr. Debra Moriarity.

Finally, I am indebted to the National Aeronautics and Space Administration for providing this unique and scientifically beneficial program.

BACKGROUND

The primary reasons for our initial interest in this research emanate from our experience in studying protein synthesis, degradation and gene expression in skeletal muscle cells, and from the extensive physical problems resulting from atrophy and weakness of skeletal and cardiac muscles following prolonged exposure to a low gravity environment (c.f., Morey-Holton and Wronski, 1981; Oganov et al., 1982; Grigor'yeva and Kozlovskaya, 1983; Leonard et al., 1983). Clearly, dramatic alterations in the balance between the rate of protein synthesis and the rate of protein degradation must accompany major changes in the quantity of muscle tissue, and it also seems exceedingly unlikely that loss of up to 25% of skeletal muscle mass could take place in the absence of switches in myofibrillar protein gene expression. The extent of skeletal muscle loss in humans and experimental animals after space flight is rather dramatic. To cite two specific examples, crew members of the first two Skylab missions maintained a negative nitrogen balance of approximately 4.5 g/day at the same time that total body potassium was also decreasing (Whedon et al., 1977). Much of this increased protein catabolism was due specifically to muscle protein degradation as evidenced by elevated rates of 3-methylhistidine excretion (Leach et al., 1979). Also, a pronounced decrease in mass and in myofibrillar cross sectional area of the soleus muscle in rats after exposure to a low gravity environment for 7 days has been reported (Goldspink et al., 1980; Riley et al., 1987).

An interesting and intriguing explanation for the effect of prolonged space flight on muscle atrophy is that a secretory defect for growth hormone (GH) may be occurring in pituitary cells (Grindeland et al., 1987a; Motter et al., 1987). Specifically, anterior pituitary cells isolated from rats flown for 7 days on the SL-3 mission secreted approximately half as much of a biologically active form of GH into culture media as ground-based controls. This observation has been qualitatively confirmed and extended by several independent approaches, including the finding that the serum concentration of GH is reduced by 50% in rats exposed to simulations of microgravity by hindlimb suspension (Motter et al., 1987).

Additionally, Grindeland et al. (1987b) have concluded that muscles of rats exposed to microgravity are significantly more resistant to exogenous and circulating GH than the skeletal muscles of control rats, since administration of GH did not alleviate muscle atrophy in animals in which it was known that GH secretion rate and serum levels of GH were also decreased by 50%. These results imply that the defect in GH utilization may extend to skeletal muscles as well.

Thus, it seems clear that exogenous factors are responsible for at least a portion of muscle atrophy; however, it also seems possible that microgravity has direct intrinsic effects on the cytoskeletal and myofibrillar contractile systems. Examples of direct effects of microgravity on cells are rather limited, but available circumstantial data are consistent with the explanation that the cytoskeletal system is involved. For example, secretory processes in general, and by definition the secretion of hormones from the pituitary, are microfilament-dependent, and it is plausible that the reduced secretion of GH described above results from a direct effect on the cytoskeletal system. Moreover, if the defect in GH utilization extends to skeletal muscle cells, and if the Insulin-like Growth Factors (IGF's, which are regulated by GH and act directly on skeletal muscle) and their receptors are internalized by cytoskeletal-dependent processes into the cytoplasm of muscle cells, then this process may also be directly affected by microgravity. Since some of the contractile proteins in the highly organized myofibrillar protein arrays in sarcomeres are nothing more than different isoforms of some of the cytoskeletal contractile proteins, an effect on expression of one class of these proteins could logically be expected to have an effect on expression of the other. Further substantiation of this possibility results from the fact that the myosin heavy chain genes are members of a rather large multigene family that may have up to twenty members in some species such as chickens, and which is known to exhibit a significant level of plasticity in its ability to have different isoforms expressed under different tissue, developmental and environmental conditions. Most crucial, however, is the fact that neither the effect of actual microgravity nor the effect of simulated microgravity on the organization, synthesis/degradation or gene expression of the contractile proteins has ever been evaluated under critically controlled conditions. One of the only possible ways to evaluate this possibility in the absence of prolonged space flights is with cultured muscle cells grown within the fibers of a rotating Clinostat. As discussed below, this instrument has the effect of mimicing a low gravity environment, and therefore allows some of the above processes to be evaluated.

OBJECTIVES

The general objective of this research was to assess the effects of exposure to simulated microgravity using a rotating Clinostat on morphological aspects of the contractile system in chicken skeletal muscle cells. Specifically, cell morphology, fusion, and patterns of contractile filament organization in skeletal muscle cell cultures grown in hollow fibers of a rotating Clinostat were evaluated. Primary techniques for this work were light microscopy and electron microscopy. Several different cell types were examined in preliminary experiments to determine the best cell type for this and subsequent projects. These cell types included chicken muscle cells, chicken cardiac muscle cells, and fetal bovine skeletal muscle cells.

METHODS AND PROCEDURES

1. Chick skeletal muscle cell cultures

Thigh muscle from 12 day broiler chick embryos was removed and disaggregated into individual cells by vortexing the muscle in growth medium on a vortex mixer at maximum speed for 20-30 seconds (Young et al., 1981). The suspension was then filtered through nylon mesh to remove connective tissue and bone, and the cells were recovered by centrifugation. Following resuspension in an appropriate amount of growth medium (Eagle's Minimum Essential Medium containing 5% chick embryo extract, 10% horse serum, 50 units/ml penicillin, 50 ug/ml streptomycin, 2.5 ug/ml fungizone) to give a concentration of 1.5×10^{10} cells/ml, the cell suspension was injected into a 70 mm long piece of 0.5 mm (inner diameter) XM-80 hollow fiber using a 1cc syringe and a 26 3/8 gauge needle. Both ends of the fiber were sealed with hot wax, and the fiber was loaded into a glass tube containing 5 ml of complete media. The fiber was held taut by a spring so that it would always be at the exact center of rotation of the Clinostat, and the glass tube was then sealed and the entire assembly loaded into the Clinostat.

Muscle cell cultures prepared as described above proliferate, fuse and begin to synthesize myofibrillar protein within 2-3 days in culture, and under most experimental conditions we have employed so far, they attain a maximum and constant quantity of myofibrillar proteins by approximately 7 days. Because the synthesis rate and the degradation rate must be exactly equal to each other in order to maintain a constant quantity of protein at steady-state, and because perturbations in either synthesis or degradation rates will result in a net change in the quantity and/or organization pattern of myofibrillar proteins, these cells will provide an excellent model for studying the dynamics of muscle protein accumulation and loss.

2. Microscopic Evaluation of Cells Grown in the Clinostat

A. Light Microscopy

Fibers were removed from Clinostat cultures and fixed in a 5% neutral buffered formalin solution for a minimum of 24 hours. The fibers were dehydrated through a graded ethanol series of 70%, 95%

and absolute ethanol, followed by a final treatment in xylene. The fibers were soaked in hot paraffin for 4-8 hours, embedded in the paraffin in plastic molds and sectioned with a microtome. The thin sections were rehydrated through a reverse graded alcohol series (100%, 95%, 70%) and deionized water, and subsequently stained with a hematoxylin/eosin or PAS myofibril stain using standard staining procedures. Slides were viewed and photomicrographs made at 45X and 100X. These light micrographs were useful for routine monitoring of the rate of cell growth and differentiation under the different experimental treatments.

B. Transmission Electron Microscopy

Fibers containing the cell cultures from the Clinostat were fixed in 4% buffered glutaraldehyde, followed by 1% osmium tetroxide fixation. The fibers were partially dehydrated in ethanol (25% and 50%) and stained with 1% uranyl acetate. After complete dehydration through a graded ethanol series (85%, 95% and absolute) the fibers were embedded in Spurr embedding medium and thin sectioned using a microtome. Sections were viewed and photographed with a Philips Model 201C transmission electron microscope. These electron micrographs were analyzed to determine if simulated microgravity had direct ultrastructural effects on the sarcomeric and/or cytoskeletal system in muscle cells.

RESULTS

The fast rotating Clinostat used for these experiments was designed at MSFC and generously loaned to us by Dr. Robert Snyder. Briefly, the Clinostat is made up of a culture chamber which rotates about a horizontal axis, and an XM-80 hollow fiber containing cells is mounted in the center of rotation. Depending on the density difference between the particles and the liquid in which they are suspended, the particles may settle within circular trails. At high enough speed of rotation, cells become motionless with respect to the gravity vector and microgravity can be simulated (Briegleb, 1983). The simulation approaches 100% for particles in suspension where centrifugal forces and Brownian motion offset each other. In the hollow fibers containing muscle cells, however, the diameter of the fiber is small enough that the centrifugal force is only approximately $0.006 \times g$ at 100 rpm of rotation. Operation of the Clinostat in the horizontal position simulates microgravity, and operation in the vertical position serves as a control since the gravity vector is always constant in the vertical position. Additional control experiments consisting of cells in hollow fibers lying horizontally in a sealed tube are also conducted to ensure that a horizontal, non-rotating control is always available. Photographs of the Clinostat are shown in the horizontal (Figure 1) and vertical positions (Figure 2). The cells are placed inside the small hollow fiber inside the rotating chamber in the foreground in Figure 1, the ends are sealed with wax, and the fiber is held taut by the spring loaded mount. This ensures that the fiber is always held at precisely the center of rotation. A tachometer for monitoring the speed of rotation is outside the photograph.

Preliminary investigations of chicken skeletal muscle cells in the Clinostat have been carried out over the past few months, and while we have made major progress in developing the appropriate conditions, it should be made very clear that we are in the early stages of analyzing muscle cells and that all conclusions are based on a limited amount of information. Additionally, we attempted to grow several different cell types to determine the best ones for future experiments, and to see if one cell type grew better than the others. Chicken skeletal muscle cells and chick cardiac muscle cells grew best, whereas fetal bovine skeletal muscle cells grew very slowly.

Light micrographs of skeletal muscle cells stained with hematoxylin and eosin after 7 days in culture are shown in Figures 3

and 4. Figure 3 is a longitudinal section (i.e., parallel to the long axis of the hollow fiber) through the monolayer of chicken skeletal muscle cells, and Figure 4 is a cross section through the monolayer of cardiac muscle cells. Figure 5 shows that bovine skeletal muscle cells did not grow as well. In these light micrographs, the cells are confluent, and although fusion can be observed occasionally from these sectioning planes, the extent of fusion and myofibril formation is far more apparent in electron micrographs (Figure 6 and Figure 7). These micrographs in Figure 6 and Figure 7 were taken from a representative control multinucleated myotube after 6 days in a hollow fiber in the absence of rotation. Cross striations are readily apparent in most sections, and the myofibrils appear to be aligned appropriately. Moreover, portions of myotube nuclei are visible.

The micrographs shown in Figure 8 and Figure 9 were taken from myotubes that had been rotating horizontally in the Clinostat at 100 rpm for 6 days. Although the following conclusions are very tentative and are based on a limited amount of information, at least two differences between the Clinostat cells and the controls have been observed so far. First, it was more difficult to find myotubes in the Clinostat cultures than in the controls (i.e., more sections from different blocks had to be examined to find myotubes, even though the same number of muscle cells was originally placed into both sets of fibers). Second, while most of the sections from the Clinostat myotubes do have some cross striations (c.f., Figure 8), we observed that the Clinostat myofibrils seem to be more poorly organized. The sections contain abundant amounts of filamentous-appearing structures that resemble myofibrillar material; however, the extent of organization and alignment seems not to be as precise as in the controls. One possible explanation that we have not yet had time to evaluate is that a higher percentage of the sections from the Clinostat cultures are at oblique angles; however, it is also easy to find this poorly organized material in sections that also contain readily identifiable sections of intact myofibrils. Thus, we can not rule out the possibility that Clinostat cultures contain as many myofibrils, but that they are more randomly oriented in the myotubes. These are questions that can only be answered by a detailed continuation of the present research and by expanding it to include the additional parameters we have proposed to measure. Again, while these results are interesting, their preliminary nature cannot be over-emphasized.

LITERATURE CITED

- Brieglieb, W. 1983. Proceedings of Workshop on Space Biology. Cologne, Germany, March 9-11, 1983. 97-101.
- Goldspink, D.F., A.J. Morton, P. Loughna and G. Goldspink. Pflugers Archiv. 407: 333-340.
- Grindeland, R. et al., 1987a. Am. J. Physiol. 252: R209-R215.
- Grindeland, R.E., T.N. Fast, M. Vasques, T. Satyanaranyane and M. Ruder. 1987. Space Life Sciences Symposium: Three Decades of Life Science Research in Space. June 21-26, 1987. Washington, D.C., pp. 82-83.
- Grigor'yeva, L.S. and I.B. Kozlovskaya. 1983. Kosmicheskaya Biologiya I Aviakosmicheskays Meditsina, 17: 29-35.
- Leach, C.S., S.I. Altchuler and N.M. Cintron-Trevino. 1983. Medicine and Science in Sports and Exercise. 15: 432-440.
- Leonard, J.I., C.S. Leach, P.C. Rambaut. 1983. Am. J. Clin. Nutr. 38: 667-669.
- Motter, K., M. Vasques, C. Hayes, M. Kendall, G. Tietjen, W.C. Hymer and R. Grindeland. 1987. Space Life Sciences Symposium: Three Decades of Life Science Research in Space. June 21-26, 1987. Washington, D.C., pp. 52-53.
- Morey-Holton, E. and T.J. Wronski. 1981. Animal models for simulating weightlessness. The Physiologist. 24 (6): S/45-S/48
- Oganov, V.S., S.A. Skuratove, L.M. Murashko, M.A. Shirvinskaya, T. Szilagy, A. Szoor, M. Rapcsak, O. Takacs, S.S. Oganessian, and Z.S. Davtyan. 1982. Biofizika, 27: 41-46.
- Riley, D.A., S. Ellis, G.R. Slocum and F.R. Sedlak. 1987. Space Life Sciences Symposium: Three Decades of Life Science Research in Space. June 21-26, 1987. Washington, D.C., pp. 106-107.
- Whedon, G.D., L. Lutwak, P. Rambaut et al. 1977. In: Biomedical Results from Skylab, NASA SP-377. R.S. Johnston and L.F. Dietlein (Eds.). Washington, DC: National Aeronautics and

Space Administration.

Young, R.B., D.G. McConnell, C.H. Suelter, T.A. Philips. 1981.
Muscle and Nerve 4, 117-124.

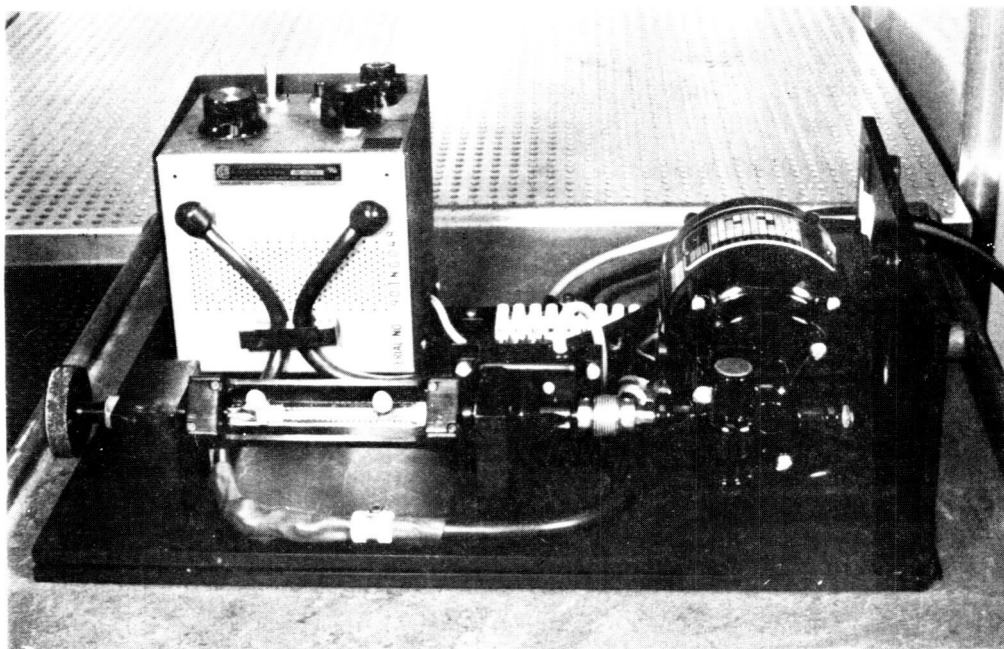


Figure 1. Clinostat: Horizontal Position

ORIGINAL PAGE IS
OF POOR QUALITY

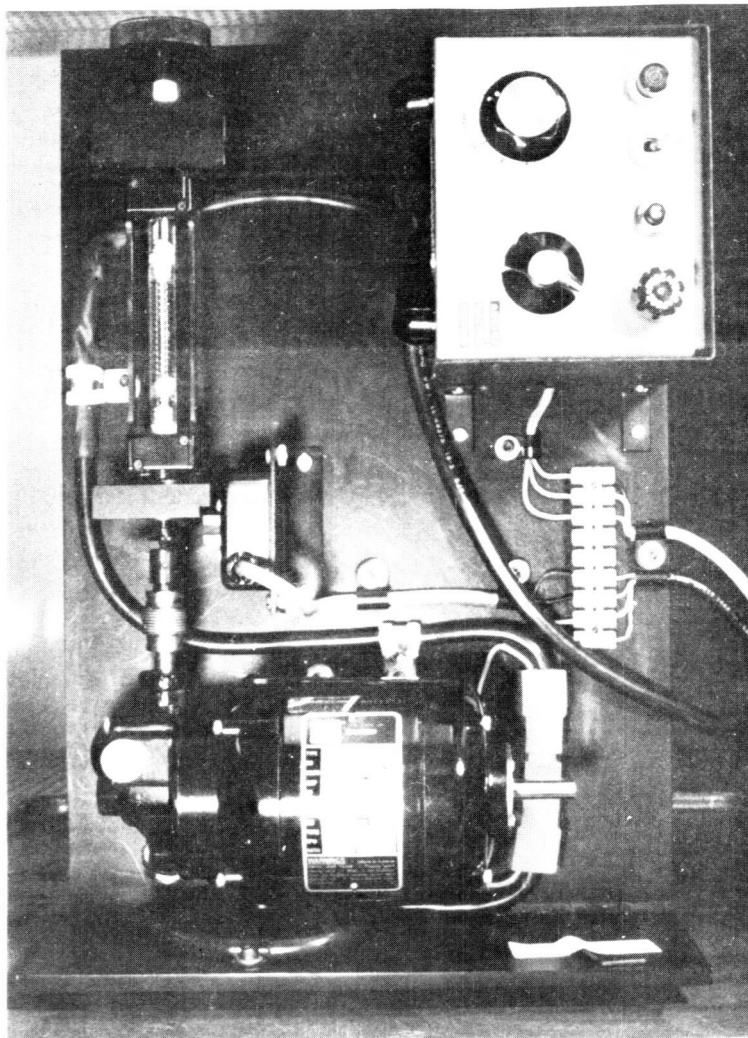


Figure 2. Clinostat: Vertical Position

ORIGINAL PAGE IS
OF POOR QUALITY

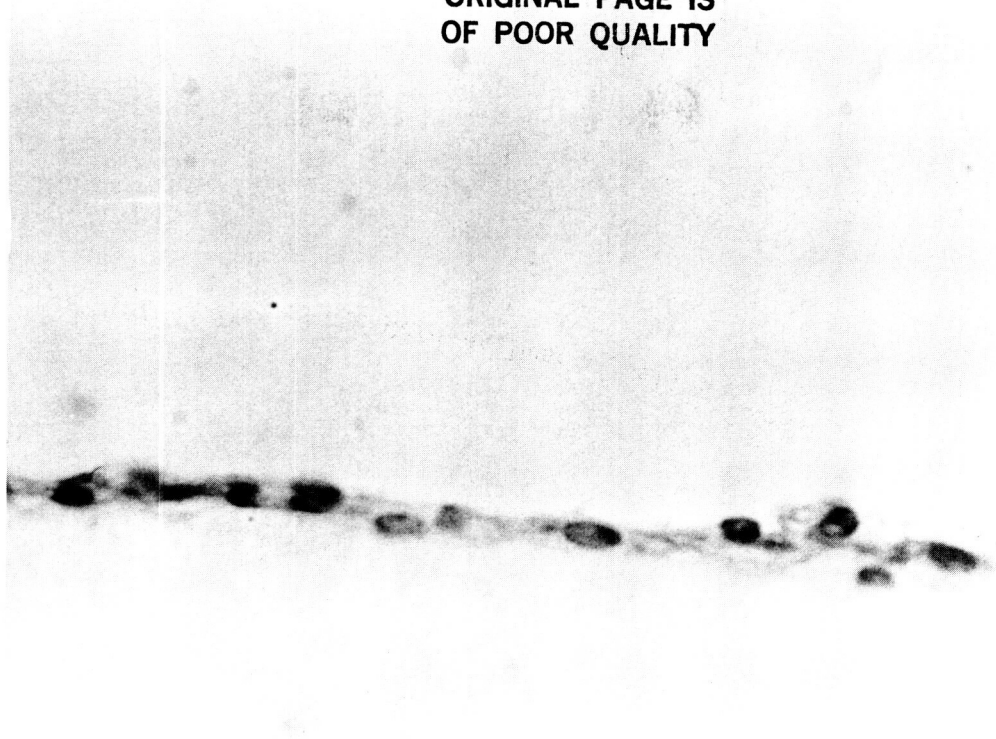


Figure 3. Chick Skeletal Muscle Cells

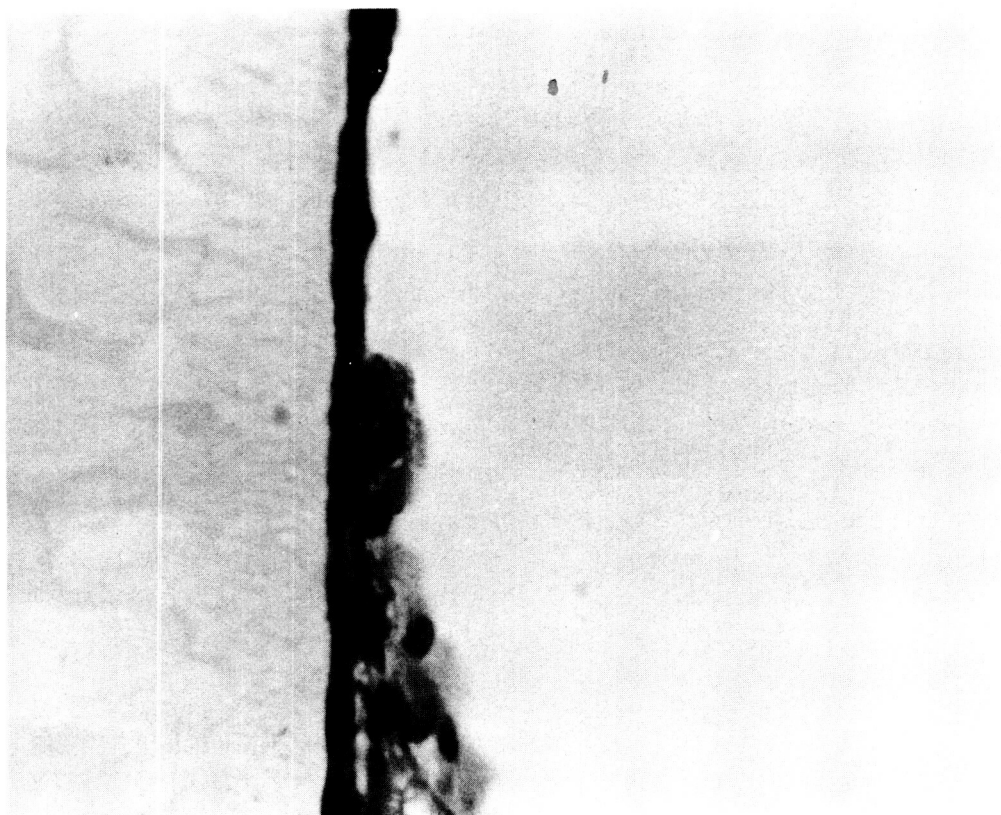


Figure 4. Chick Cardiac Muscle Cells



Figure 5. Bovine Skeltal Muscle Cells

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

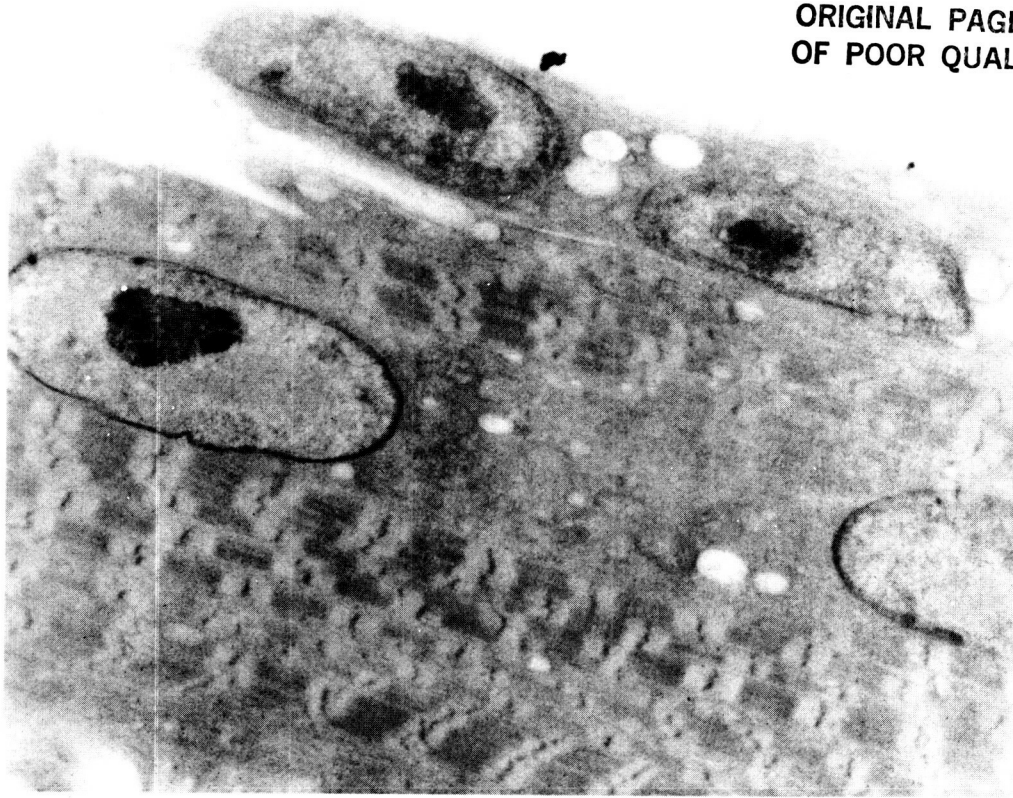


Figure 6. Electron Micrograph of Control Chick Skeletal Muscle

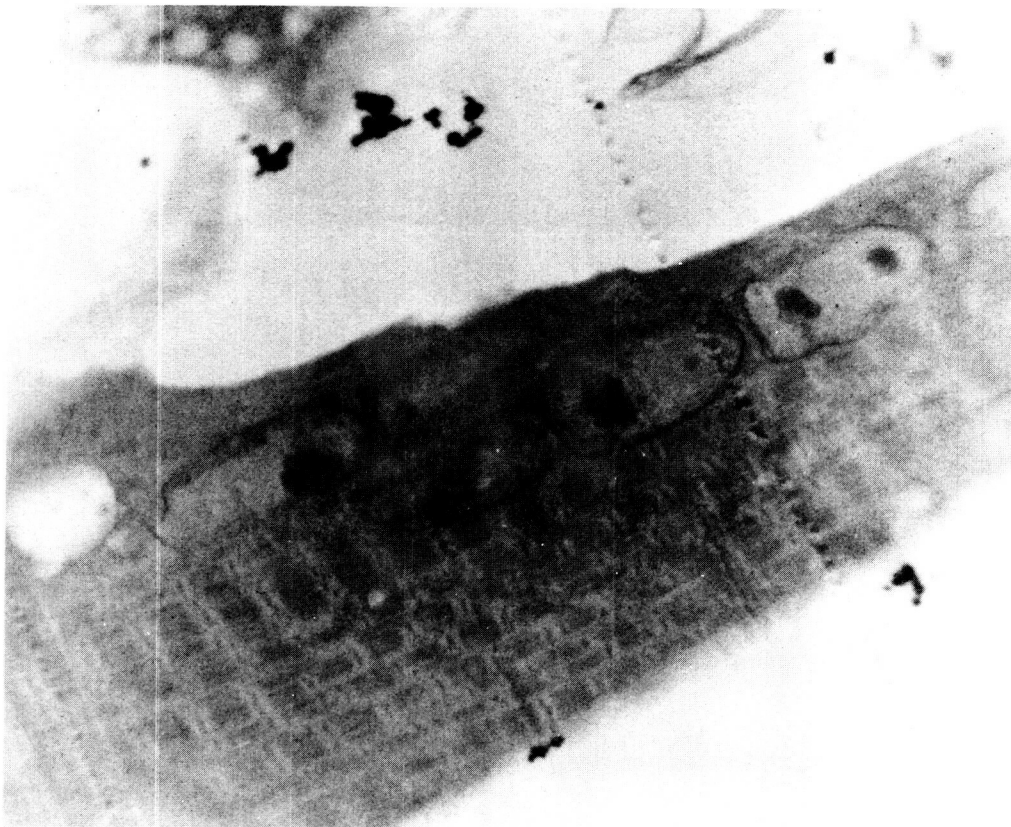


Figure 7. Electron Micrograph of Control Chick Skeletal Muscle

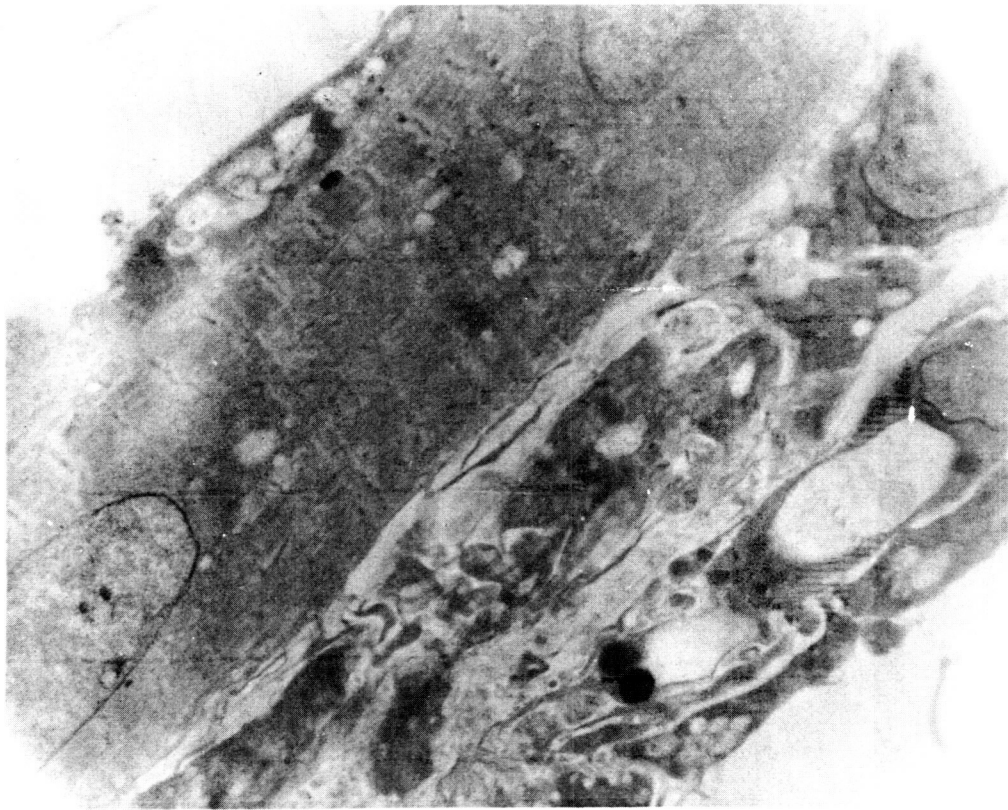


Figure 8. Electron Micrograph of Clinostat Treated Chick Skeletal Muscle Cells.

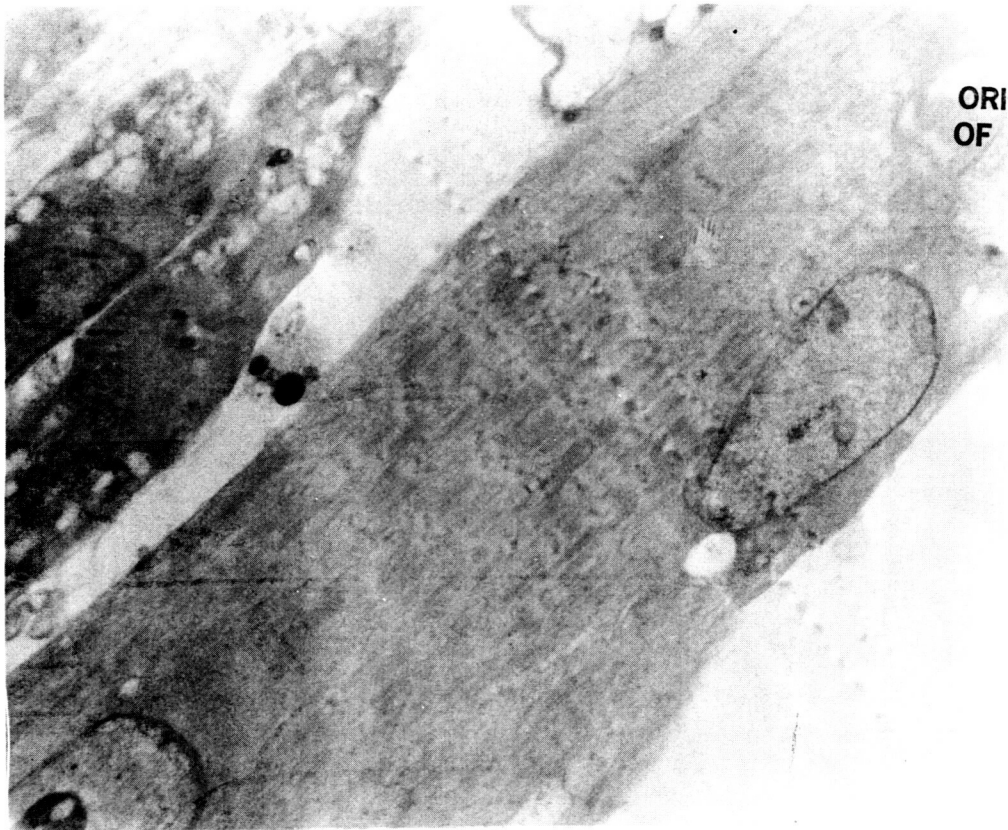


Figure 9. Electron Micrograph of Clinostat Treated Chick Skeletal Muscle Cells.

N89-21759

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

**MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA**

Rotordynamic Analysis of a Bearing Tester

Prepared by:

Richard A. Zalik

Academic Rank:

Professor

University and Department:

Auburn University

**Department of Algebra,
Combinatorics and Analysis**

NASA/MSFC

Laboratory:

Structures and Dynamics

Division:

Control Systems

Branch:

Mechanical Systems Control

MSFC Colleague:

Thomas H. Fox

Date:

August 30, 1988

Contract No.:

NGT 01-002-099

The University of Alabama

**Rotordynamic Analysis of a
Bearing Tester**

R. A. Zalik

**Auburn University
Division of Mathematics
120 Mathematics Annex Building
Auburn, AL 36849-5307**

Acknowledgement

This work was written at Marshall Space Flight Center with the support of a NASA/ASEE Faculty Fellowship. The author would like to thank NASA for the use of their facilities, and acknowledge with gratitude the assistance of several MSFC personnel, in particular P. Broussard, T. H. Fox, G. von Pragenau, S. Ryan and J. Slaby.

Objectives

The objective of this paper is to study the stability characteristics of a bearing tester. We verify our conclusions using numerical simulations of a realistic model.

Abstract. We study the properties of the solutions of a system of four coupled nonlinear differential equations that model the behavior of the rotating shaft of a bearing tester. In particular, we show how bounds for the solutions of these equation can be obtained from bounds for the solutions of the linearized equations. By studying the behavior of the Fourier transforms of the solutions, we are also able to predict the approach to the stability boundary. These conclusions are verified by means of numerical solutions of the equations, and of power spectrum density (PSD) plots.

1. Introduction.

In this study we continue the investigation of the properties of the solutions of mathematical models of rotating machinery initiated by Day [1]. Both Day and this author [2] have studied the behavior of a simple Jeffcott model with deadband, viz. a system of coupled differential equations that represent the behavior of a rotating shaft.

The purpose of this paper is to examine the properties of the solutions of a model of a bearing tester. This is a device designed to estimate the life expectancy of bearings under realistic conditions of loads and acceleration in cryogenic fluids. Our study will help determine safety margins for its operation.

We consider a bearing tester with two seals and two bearings with deadband. A sketch of this mechanical system can be seen in Fig. 1.

2. General Theory

2.1 Derivation of the Bearing Tester Equations.

We assume that the shaft is rotating with angular velocity ω along an axis close to the x -axis, that both bearings are at the same distance a from the center of symmetry of the shaft, that both seals are at the same distance b from this center of symmetry, and that the shaft cannot move in the direction of the x -axis. We also assume that both seals have the same damping C_s , stiffness K_s , and cross coupling stiffness Q_s , and that both bearings have the same stiffness K_b .

For $j = 1, 2$, let δ_j denote the magnitude of the deadband at bearing j ; let v_{yj} and v_{zj} describe the displacement of the center of the shaft at bearing j , and let w_{yj} and w_{zj} be similarly defined for the seals (see Fig. 2). Let m denote the mass of the shaft. If $r_j = (v_{yj}^2 + v_{zj}^2)^{1/2}$,

$h_j(t) = 1$ if $r_j \leq \delta_j$, and $h_j(t) = \delta_j/r_j$ if $r_j > \delta_j$, then the equations that describe the movement of the shaft are the following:

$$K_b[1-h_1(t)]v_{y1} + K_b[1-h_2(t)]v_{y2} + K_s[w_{y1} + w_{y2}] + Q_s[w_{z1} + w_{z2}] + C_s[w'_{y1} + w'_{y2}] + (m/2)[v''_{y1} + v''_{y2}] = g_1(t) \quad (1)$$

$$K_b[1-h_1(t)]v_{z1} + K_b[1-h_2(t)]v_{z2} + K_s[w_{z1} + w_{z2}] - Q_s[w_{y1} + w_{y2}] + C_s[w'_{z1} + w'_{z2}] + (m/2)[v''_{z1} + v''_{z2}] = g_2(t) \quad (2)$$

$$-aK_b[1-h_1(t)]v_{y1} + aK_b[1-h_2(t)]v_{y2} + bK_s[w_{y2} - w_{y1}] + bQ_s[w_{z2} - w_{z1}] + bC_s[w'_{y2} - w'_{y1}] + (I_2/2a)[v''_{y2} - v''_{y1}] - (\omega I_1/2a)[v'_{z2} - v'_{z1}] = m_1(t) \quad (3)$$

$$-aK_b[1-h_1(t)]v_{z1} + aK_b[1-h_2(t)]v_{z2} + bK_s[w_{z2} - w_{z1}] - bQ_s[w_{y2} - w_{y1}] + bC_s[w'_{z2} - w'_{z1}] + (I_2/2a)[v''_{z2} - v''_{z1}] + (\omega I_1/2a)[v'_{y2} - v'_{y1}] = m_2(t) \quad (4)$$

$$v_{y1} + v_{y2} = w_{y1} + w_{y2}, \quad v_{z1} + v_{z2} = w_{z1} + w_{z2} \quad (5)$$

$$v_{y1} - v_{y2} = w_{y1} - w_{y2}, \quad v_{z1} - v_{z2} = w_{z1} - w_{z2}, \quad (6)$$

where I_1 is the axial inertia, I_2 is the rotational inertia about the axis transversal to the shaft, and $g_1(t)$, $g_2(t)$, $m_1(t)$, $m_2(t)$ are given as follows:

$$g_1(t) = \omega^2 m[(e_{y1} + e_{y2})\cos \omega t - (e_{z1} + e_{z2})\sin \omega t]$$

$$g_2(t) = \omega^2 m[(e_{z1} + e_{z2})\cos \omega t + (e_{y1} + e_{y2})\sin \omega t]$$

$$m_1(t) = \omega^2 m[(e_{y1} - e_{y2})\cos \omega t - (e_{z1} - e_{z2})\sin \omega t]$$

$$m_2(t) = \omega^2 m[(e_{z1} - e_{z2})\cos \omega t + (e_{y1} - e_{y2})\sin \omega t],$$

where e_{y1} , e_{y2} , e_{z1} and e_{z2} represent the mass imbalance. In our analyses we shall assume that g_1 , g_2 , m_1 , and m_2 are arbitrary continuous and bounded functions.

Note that (1) and (2) are force equations, (3) and (4) are moment equations, and (5) and (6) are derived from the symmetry assumptions on bearings and seals. Setting $v_j = v_{yj} + iv_{zj}$, $w_j = w_{yj} + iw_{zj}$, $g(t) = g_1(t) + ig_2(t)$ and $m(t) = m_1(t) + im_2(t)$, we obtain:

$$K_b[1-h_1(t)]v_1 + K_b[1-h_2(t)]v_2 + (K_s - iQ_s)(w_1 + w_2) + C_s[w_1' + w_2'] + (m/2)[v_1'' + v_2''] = g(t). \quad (7)$$

$$-aK_b[1-h_1(t)]v_1 + aK_b[1-h_2(t)]v_2 + b(K_s - iQ_s)(w_2 - w_1) + bC_s[w_2' - w_1'] + (I_2/2a)[v_2'' - v_1''] + i(\omega I_1/2a)[v_2' - v_1'] = m(t), \quad (8)$$

and

$$v_1 + v_2 = w_1 + w_2, \quad v_1 - v_2 = w_1 - w_2 \quad (9)$$

In view of (9), if we set $v = v_1 + v_2$, $u = v_2 - v_1$, $q_1 = h_1 v_1$, and $q_2 = h_2 v_2$, (7) and (8) can be written in the following form:

$$(m/2)v'' + C_s v' + (K_b + K_s - iQ_s)v - K_b[q_1 + q_2] = g(t),$$

and

$$(I_2/2a)u'' + (bC_s + i\omega I_1/2a)u' + (aK_b + bK_s - ibQ_s)u + aK_b[q_1 - q_2] = m(t).$$

Thus, setting $C_1 = 2C_s/m$, $C_2 = (2abC_s + i\omega I_1)/I_2$, $K_1 = 2K_b/m$, $K_2 = (2a^2K_b)/I_2$, $A_1 = 2K_s/m$, $A_2 = 2abK_s/I_2$, $B_1 = 2Q_s/m$, $B_2 = 2abQ_s/I_2$, $M_1 = A_1 + K_1 - iB_1$,

$M_2 = A_2 + K_2 - 1B_2$, $f_1(t) = (2/m) g(t)$, $f_2(t) = (2a/I_2)m(t)$, and

$$p_1 = q_1 + q_2, \quad p_2 = q_1 - q_2, \quad (10)$$

we finally obtain the Bearing Tester equations:

$$v'' + C_1 v' + M_1 v - K_1 p_1 = f_1(t), \quad (11)$$

and

$$u'' + C_2 u' + M_2 u + K_2 p_2 = f_2(t). \quad (12)$$

We shall assume that $B_1, B_2, C_1, C_2, K_1, K_2, \delta_1, \delta_2$ are positive, and A_1, A_2 and t are nonnegative.

Since $v_1 = (1/2)(v - u)$ and $v_2 = (1/2)(v + u)$, $p_1(t)$ and $p_2(t)$ can be expressed in terms of v and u using (10) and the following representations for $q_1(t)$ and $q_2(t)$:

$$q_1(t) = \begin{cases} (v-u)/2, & \text{if } |v-u| \leq 2\delta_1, \\ \delta_1(v-u)/|v-u|, & \text{if } |v-u| > 2\delta_1, \end{cases} \quad (13)$$

and

$$q_2(t) = \begin{cases} (v+u)/2, & \text{if } |v+u| \leq 2\delta_2, \\ \delta_2(v+u)/|v+u|, & \text{if } |v+u| > 2\delta_2. \end{cases} \quad (14)$$

2.2 Existence, uniqueness, and representation formulas.

We have transformed the system of equations (1) - (6) into the equivalent system (11), (12). This is a system of coupled nonlinear differential equations similar to the Jeffcott equations we studied in [2]. The existence and uniqueness of their solutions (and therefore, of the solutions of the original system), follow by the same argument employed for the Jeffcott equations, and need not be repeated here.

$$\begin{aligned} \text{Let } Q_1 &= C_1^2 - 4(A_1 + K_1), \\ \beta_1 &= 8^{-1/2} [-Q_1 + (Q_1^2 + 16B_1^2)^{1/2}]^{1/2}, \end{aligned} \quad (15)$$

$$\alpha_1 = [\beta_1^{-1} B_1 - C_1]/2, \quad \alpha_1' = -[\beta_1^{-1} B_1 + C_1]/2, \quad (16)$$

$$\lambda_1 = \alpha_1 + i\beta_1, \quad \lambda_2 = \alpha_1' - i\beta_1. \quad (17)$$

Then, as in [2], it is readily seen that λ_1 and λ_2 are the solutions of the characteristic equation $\lambda^2 + C_1\lambda + M_1 = 0$, and therefore

$$v_h = c_1 \exp(\lambda_1 t) + c_2 \exp(\lambda_2 t) \quad (18)$$

is the general solution of

$$v'' + C_1 v' + M_1 v = 0, \quad (19)$$

Similarly, if $\gamma_1 = \alpha_2 + i\beta_2$ and $\gamma_2 = \alpha_2' + i\beta_2'$ are the solutions of the characteristic equation $\gamma^2 + C_2\gamma + M_2 = 0$, it is clear that

$$u_h = d_1 \exp(\gamma_1 t) + d_2 \exp(\gamma_2 t) \quad (20)$$

is the general solution of

$$u'' + C_2 u' + M_2 u = 0. \quad (21)$$

Without loss of generality, we shall always assume that $\alpha_2' \leq \alpha_2$. If C_2 is real (i.e., if $\omega I_1 = 0$), then formulas similar to (15), (16) and (17) obtain for γ_1 and γ_2 .

If v_p and u_p are particular solutions of the linearized Bearing Tester equations

$$v'' + C_1 v' + M_1 v = f_1(t) \quad (22)$$

and

$$u'' + C_2 u' + M_2 u = f_2(t), \quad (23)$$

then, setting $v_\ell = v_h + v_p$, $u_\ell = u_h + u_p$,

$$G_1(t) = (\lambda_1 - \lambda_2)^{-1} [\exp(\lambda_1 t) - \exp(\lambda_2 t)],$$

$$G_2(t) = (\gamma_1 - \gamma_2)^{-1} [\exp(\gamma_1 t) - \exp(\gamma_2 t)],$$

and proceeding as in [2], we readily deduce that (11) and (12) are equivalent to the following nonlinear Volterra integral equations of convolution type:

$$v(t) = v_\ell(t) + P_1(t) \quad (24)$$

and

$$u(t) = u_1(t) + P_2(t), \quad (25)$$

where the perturbation terms $P_j(t)$ are given by:

$$P_j(t) = K_j \int_0^t G_j(t-x) p_j(x) dx, \quad j = 1, 2. \quad (26)$$

Thus, the existence and uniqueness of the solutions of (11) and (12) also follow from the existence and uniqueness of the solutions of (24) and (25) (cf. e.g. [3], [4]).

2.2 Bounds

Let $D_j(t) = \int_0^t |G_j(t-x)| dx$, and let $\delta = \delta_1 + \delta_2$. Since (10), (11), (13) and (14) imply that

$$|p_j(t)| \leq \delta, \quad j = 1, 2, \quad (27)$$

we readily conclude that

$$|P_j(t)| \leq \delta K_j D_j(t), \quad j = 1, 2 \quad (28)$$

Let $D_1 = |\lambda_1 - \lambda_2|^{-1} (|\alpha_1|^{-1} + |\alpha_1'|^{-1})$, $D_2 = |\gamma_1 - \gamma_2|^{-1} (|\alpha_2|^{-1} + |\alpha_2'|^{-1})$.

Note that $\alpha_j' \leq \alpha_j$. Thus if $\alpha_j < 0$ we readily see that D_j is a steady state bound for $D_j(t)$. From these inequalities we derive, as in [2], the following conclusions:

1. If $\alpha_1 < 0$ and $|v_p| \leq M_1$, then the steady state solution v_∞ of (11) satisfies the following inequality:

$$|v_\infty| \leq M_1 + \delta K_1 D_1$$

whereas if $\alpha_2 < 0$ and $|u_p| \leq M_2$, the steady state solution u_∞ of (12) satisfies the inequality

$$|u_\infty| \leq M_2 + \delta K_2 D_2$$

2. If $\alpha_j = 0$, the perturbation term $P_j(t)$ can grow at most linearly.
3. If $\alpha_j > 0$, the order of growth of $P_j(t)$ cannot exceed $\exp(\alpha_j t)$; note that the order or magnitude of all nonzero solutions of (19) or

(21) cannot exceed $\exp(\alpha_j t)$.

Since our assumptions imply that $f_1(t)$ and $f_2(t)$ are bounded, we have therefore shown that the study of the boundedness of the solutions of (11) or (12) reduces to the study of the boundedness of the solutions of (19) or (21).

If $\alpha_1 < 0$ and $\alpha_2 < 0$ we shall say that the system (11), (12) (or (1)-(6)) is stable, if $\alpha_1 = 0$ and $\alpha_2 \leq 0$, or $\alpha_1 \leq 0$ and $\alpha_2 = 0$, that the system has reached the stability boundary, and if $\alpha_1 > 0$ or $\alpha_2 > 0$, that the system is unstable. Thus the system is stable if all its solutions are bounded.

2.3. Estimates for β_1 and β_2

We obtain estimates for the β_j in terms of the coefficients of (11) and (12) and the signs of the α_j . These estimates yield a simple method for determining the stability of the system. Since (19) is identical with [2, (7)], we know the following:

1. If $\alpha_1 < 0$, then $B_1/C_1 < \beta_1 < (A_1 + K_1)^{1/2}$, and $\alpha_1' < 0$.
2. If $\alpha_1 = 0$, then $B_1/C_1 = \beta_1 = (A_1 + K_1)^{1/2}$, and $\alpha_1' < 0$.
3. If $\alpha_1 > 0$, then $(A_1 + K_1)^{1/2} < \beta_1 < B_1/C_1$.

If C_2 is real (i.e., if $\omega I_1 = 0$), we also have:

4. If $\alpha_2 < 0$, then $B_2/C_2 < \beta_2 < (A_2 + K_2)^{1/2}$, and $\alpha_2' < 0$.
5. If $\alpha_2 = 0$, then $B_2/C_2 = \beta_2 = (A_2 + K_2)^{1/2}$, and $\alpha_2' < 0$.
6. If $\alpha_2 > 0$, then $(A_2 + K_2)^{1/2} < \beta_2 < B_2/C_2$.

From these conclusions we also infer that if $f_1(t)$ and $f_2(t)$ are bounded, and $I_1 = 0$, then the system (1) - (6) is stable if and only if

$$B_1/C_1 < (A_1 + K_1)^{1/2}, \text{ and } B_2/C_2 < (A_2 + K_2)^{1/2}.$$

2.4 Resonance

Proceeding as in [2], we readily see that if $f_j(t) = A_j \exp(i\omega t)$, $j = 1, 2$, then (1) - (6) can be in resonance only if $\alpha = 0$ or $Q_s/C_s = [2a(bK_s + aK_b)/(I_1 + I_2)]^{1/2}$.

3. Harmonic Analysis of the solutions.

3.1. Preliminaries.

We now study the properties of the Fourier transforms of the solutions.

Following standard practice, we consider a time interval of the form (c, d) $0 \leq c < d < \infty$. Let $g^{(c,d)}(t) = g(t)$ if $c \leq t \leq d$, and let $g^{(c,d)}(t) = 0$ otherwise. Thus, if F denotes the Fourier transform operator, we have:

$$F[g^{(c,d)}](s) = (2\pi)^{-1/2} \int_c^d g(t) \exp(ist) dt,$$

and proceeding as in [2] we see that

$$\lim_{c \rightarrow \infty} F[v_h^{(c,d)}](s) = 0, \text{ and } \lim_{c \rightarrow \infty} F[u_h^{(c,d)}](s) = 0, \quad (29)$$

and therefore

$$\lim_{c \rightarrow \infty} F[v_1^{(c,d)}](s) = 0, \text{ and } \lim_{c \rightarrow \infty} F[v_2^{(c,d)}](s) = 0, \quad (30)$$

We want to study the properties of the graphs of the absolute values of $F[v^{(c,d)}](s)$ and $F[u^{(c,d)}](s)$. From (29) it is clear that in order to obtain useful information we need to study the Fourier transforms of the perturbation terms $P_j(t)$.

3.2. Analysis of the perturbation terms.

$$\text{Let } Q_1(t) = \int_0^t \exp[\lambda_1(t-x)] p_1(x) dx, \text{ and}$$

$R_2(t) = \int_0^t \exp[\lambda_2(t-x)] p_1(x) dx$, and let $Q_2(t)$, $R_2(t)$ be similarly defined in terms of γ_1 , γ_2 , and $p_2(x)$. Clearly

$$p_1^{(c,d)}(t) = K_1(\lambda_1 - \lambda_2)^{-1} [Q_1^{(c,d)}(t) - R_1^{(c,d)}(t)].$$

We have:

$$\begin{aligned} F[Q_1^{(c,d)}] &= (2\pi)^{-1/2} \int_c^d \int_0^t \exp[\lambda_1(t-x)] p_1(x) dx \exp(-st) dx \\ &= (2\pi)^{-1/2} \int_c^d \int_0^t \exp(-\lambda_1 x) p_1(x) dx \exp[t(\lambda_1 - s)] dt, \end{aligned}$$

and integrating by parts we obtain:

$$\begin{aligned} F[Q_1^{(c,d)}](s) &= \\ &= (2\pi)^{-1/2} (\lambda_1 - s)^{-1/2} [\exp(-ds) p_1(d) - \exp(-cs) p_1(c)] - \int_c^d \exp(-s(i)) p_1(t) dt \\ &= M_1(s, c, d) / (\lambda_1 - s), \end{aligned}$$

where, since $|p_1(t)| \leq \delta$,

$$|M_1(s, c, d)| \leq (2\pi)^{-1/2} (2 + d - c) \delta. \quad (31)$$

Using the same argument we also see that

$$F[R_1^{(c,d)}](s) = M_2(s, c, d) / (\lambda_2 - s),$$

where $M_2(s, c, d)$ satisfies an inequality similar to (31). Thus,

$$F[p_1^{(c,d)}](s) = K_1(\lambda_1 - \lambda_2)^{-1} [M_1(s, c, d) / (\lambda_1 - s) - M_2(s, c, d) / (\lambda_2 - s)]. \quad (32)$$

Similarly,

$$F[p_2^{(c,d)}](s) = K_2(\gamma_1 - \gamma_2)^{-1} [M_3(s, c, d) / (\gamma_1 - s) - M_4(s, c, d) / (\gamma_2 - s)], \quad (33)$$

where $|M_3(s, c, d)|$ and $|M_4(s, c, d)|$ are bounded by $(2\pi)^{-1/2} (2 + d - c) \delta$.

Note, moreover, that there is no reason why $M_1(s, c, d)$ or $M_3(s, c, d)$ should vanish as $c \rightarrow \infty$, provided we keep the difference $d - c$ constant.

Since $v_1 = (1/2)(v - u)$ and $v_2 = (1/2)(v + u)$, for $j = 1, 2$ we obtain:

$$v_j = T_j + L_j + H_j,$$

where L_j is a particular solution of (7) or (8), $\lim_{c \rightarrow \infty} F[T_j^{(c,d)}](s) = 0$,

$$F[H_1^{(c,d)}](s) = (1/2)(F[p_1^{(c,d)}](s) - F[p_2^{(c,d)}](s)), \quad (34)$$

and

$$F[H_2^{(c,d)}](s) = (1/2)(F[p_1^{(c,d)}](s) + F[p_2^{(c,d)}](s)). \quad (35)$$

Since $\lambda_1 - si = \alpha_1 + (\beta_1 - si)$, and $\gamma_1 - si = \alpha_2 + (\beta_2 - si)$, this means that the only value for which $F[H_1^{(c,d)}](s)$ and $F[H_2^{(c,d)}](s)$ may diverge as $\alpha_1 \rightarrow 0^-$ is $s = \beta_1$, and the only value for which they may diverge as $\alpha_2 \rightarrow 0^-$ is $s = \beta_2$.

Let $\sigma = \max \{\alpha_1, \alpha_2\}$. We identify the nonlinear natural frequency with those frequencies at which the PSD plots have relative maxima that become unbounded $\sigma \rightarrow 0^-$. The method we have used to reach the conclusions of this section is much simpler than the one we employed in [2].

3.3.3 Conclusions

If $\sigma = \max \{\alpha_1, \alpha_2\}$ and $\xi = \beta$ if $\sigma = \alpha$ or $\xi = \beta_2$ otherwise, we conclude that as $\sigma \rightarrow 0$ $F[v_1^{(c,d)}](s)$ and $F[v_2^{(c,d)}](s)$ may diverge only at $s = \xi$ (or, if both $\alpha_1 \rightarrow 0^-$ and $\alpha_2 \rightarrow 0^-$ simultaneously, at both β and β'), that for σ negative and constant, but sufficiently close to zero, the graphs of the absolute values of $F[v_j^{(c,d)}](s)$, $j = 1, 2$ will have spikes near $s = \xi$ (or, in PSD plots, near $\xi/s\pi$), and that the magnitudes of these spikes need not decrease with time (i.e., as $c \rightarrow \infty$).

4. Examples

We now study the behavior of the solutions of (11), (12) for various rotating speeds. We assume that $I_1 = 0$, $I_2 = 1.65 \text{ lbf.in.sec}^2$, $\delta = 0.0015 \text{ in.}$, $m = 0.0587 \text{ lb.sec}^2/\text{in.}$, $a = 4.15 \text{ in.}$, $b = 7.1 \text{ in.}$ If ϕ denotes the angular

speed of the shaft in Hz., then $K_s = 0.0398\phi$ lb/in, $Q_s = 0.0201\phi$ lb/in, $C_s = 0.00001361\phi$ lb/in, and $K_b = -1,800\phi + 105,480,000$. We also assume that $e_{y1} = 5.7 \times 10^{-6}$ lb.sec²/in, $e_{y2} = -e_{y1}$, $e_{z1} = e_{z2} = 0$, $v(0) = u(0) = \delta$, and $v'(0) = u'(0) = 1$. These values have been obtained from an actual bearing tester. Since, as we have already shown, when $\alpha_1 = 0$ we have $B_1/C_1 = (A_1 + K_1)^{1/2}$, we readily see that the value of ϕ for which this happens is $\phi_{c1} = 939$ Hz. = 56,359 rpm, and that $\alpha_1 < 0$ if $\phi < \phi_{c1}$. Since $I_1 = 0$, we can apply a similar procedure to conclude that $\phi_{c2} = 916$ Hz. = 54,950 rpm, and that $\alpha_2 < 0$ if $\phi < \phi_{c2}$. If f_1 denotes the frequency that corresponds to the value of β_1 when $\alpha_1 = 0$, it is readily seen that $\phi_1 = 245$ Hz. If f_2 denotes the frequency that corresponds to β_2 when $\alpha_2 = 0$, we see that also $\phi_2 = 245$ Hz. Since α_2 vanishes before α_1 , we reach the stability boundary when the shaft's rotating speed is 54,950 rpm.

In Figs. 3 through 8 we see PSD plots for v_2 for various values of ϕ ranging from 30,000 rpm to 57,000 rpm. (The plots for v_1 are similar). To obtain these plots we first solved (11), (12) using a fourth order Runge-Kutta method. We then applied a Fast Fourier algorithm. The plots were obtained using 256 points and linear interpolation, and are for the time interval 5.120 sec. $< t < 5.632$ sec. The frequencies are measured in Hz.

Since the mass imbalance is so small, the forcing frequency ϕ is

undetectable. We see that as ϕ increases, the location of the nonlinear natural frequency σ remains at 246 Hz. The magnitude of the spike increases steadily, until around 50,000 rpm it starts to climb steeply. These examples show that the nonlinear natural frequency may appear well before the stability boundary is reached. They also show that the location of this frequency is not a good indicator of stability margins, and that the approach to the stability boundary is accompanied by a steep increase in the size of the spike. These conclusions are similar to those we reached for a simple Jeffcott model in [2].

References

1. W. B. Day, Asymptotic Expansions in Nonlinear Rotordynamics, Quart. Appl. Math. 44 (1987), 779-792.
2. R. A. Zalik, The Jeffcott Equations in Nonlinear Rotordynamics, Quart. Appl. Math., (to appear).
3. T. L. Saaty, "Modern Nonlinear Equations", Dover, New York, 1981.
4. F. G. Tricomi, "Integral Equations", Interscience, New York, 1957.

FIGURES

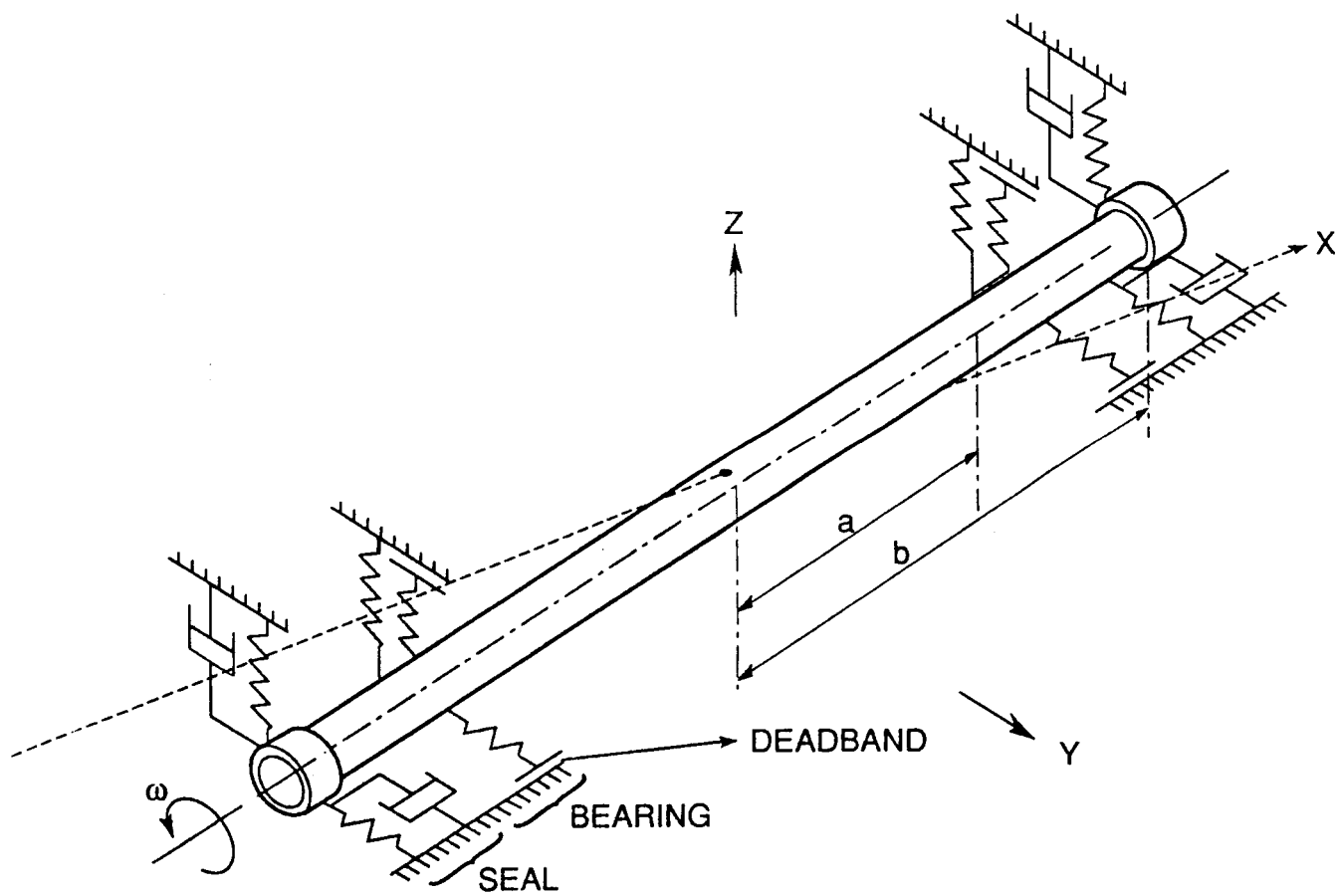


FIGURE 1

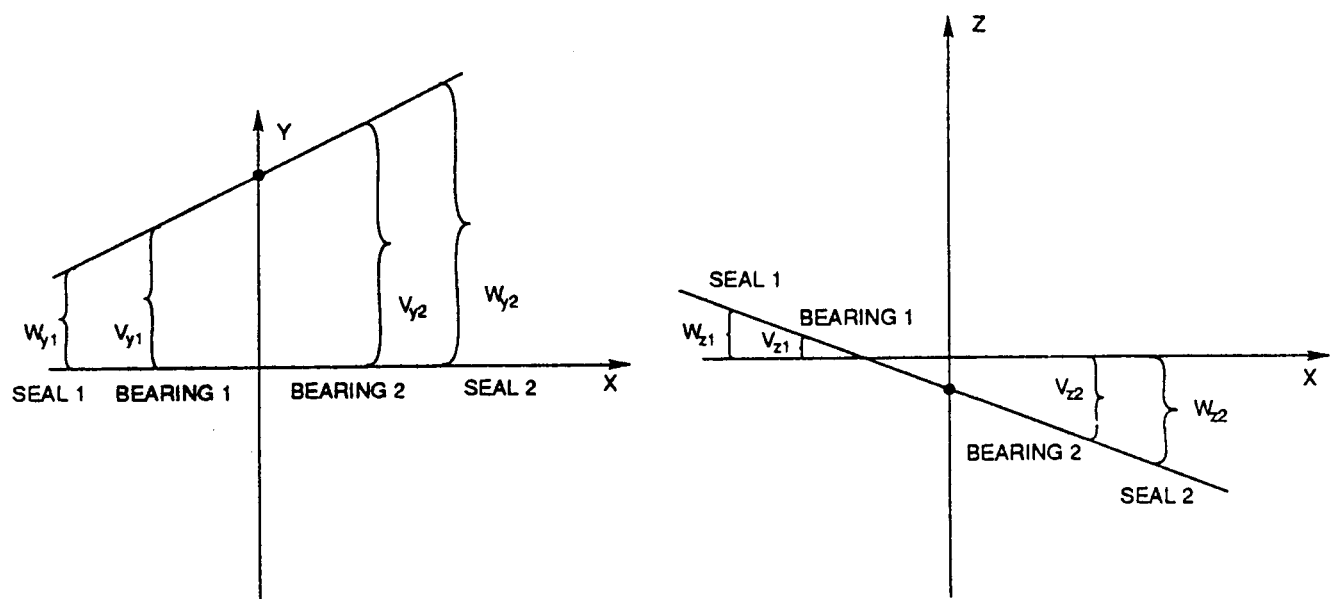
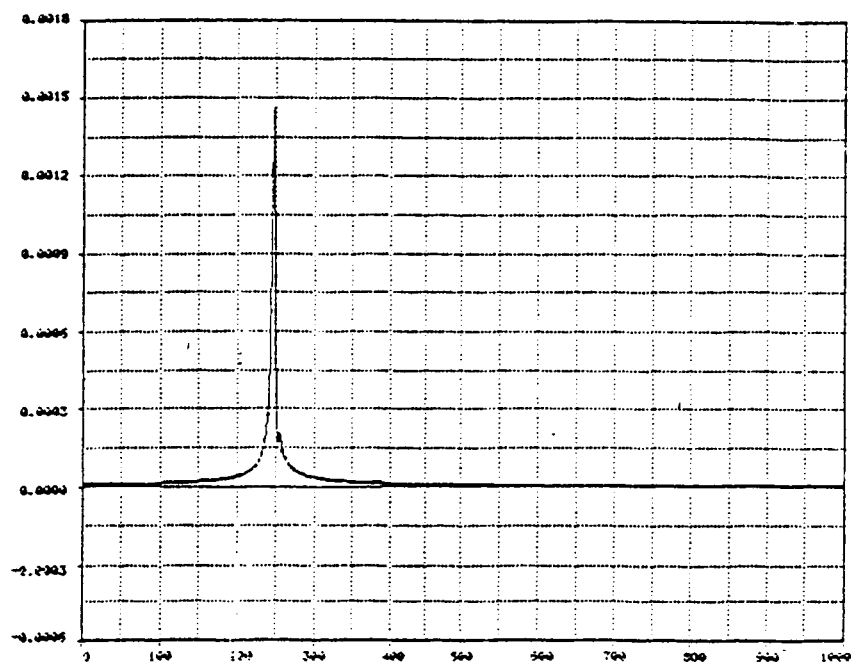


FIGURE 2

ORIGINAL PAGE IS
OF POOR QUALITY

30000 RPM



40000 RPM

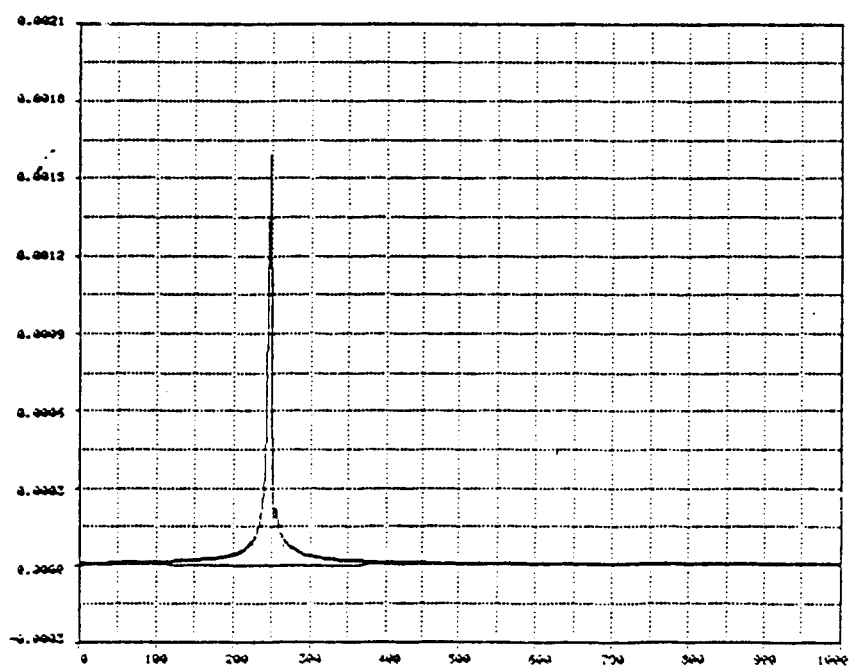


Figure 3

ORIGINAL PAGE IS
OF POOR QUALITY

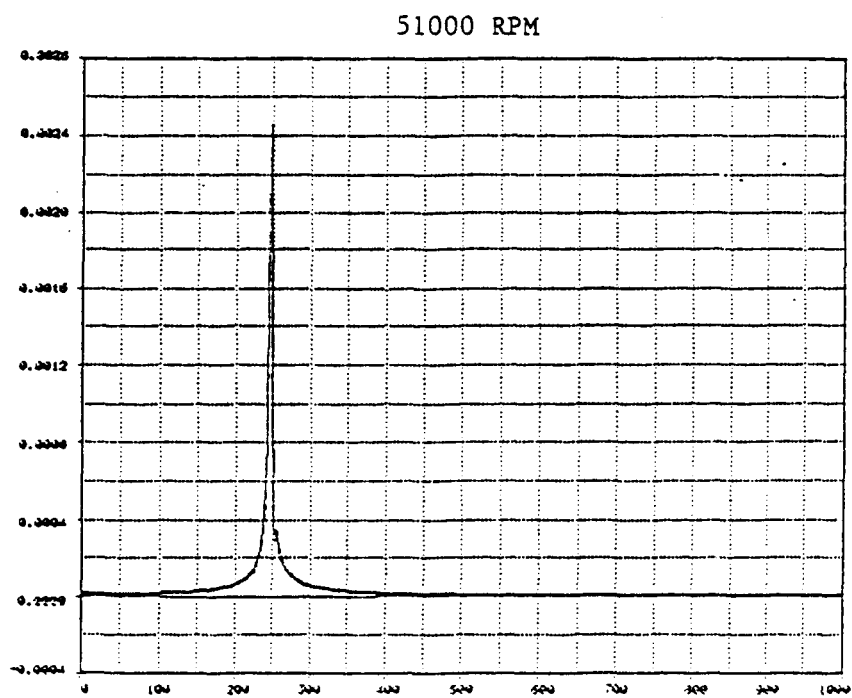
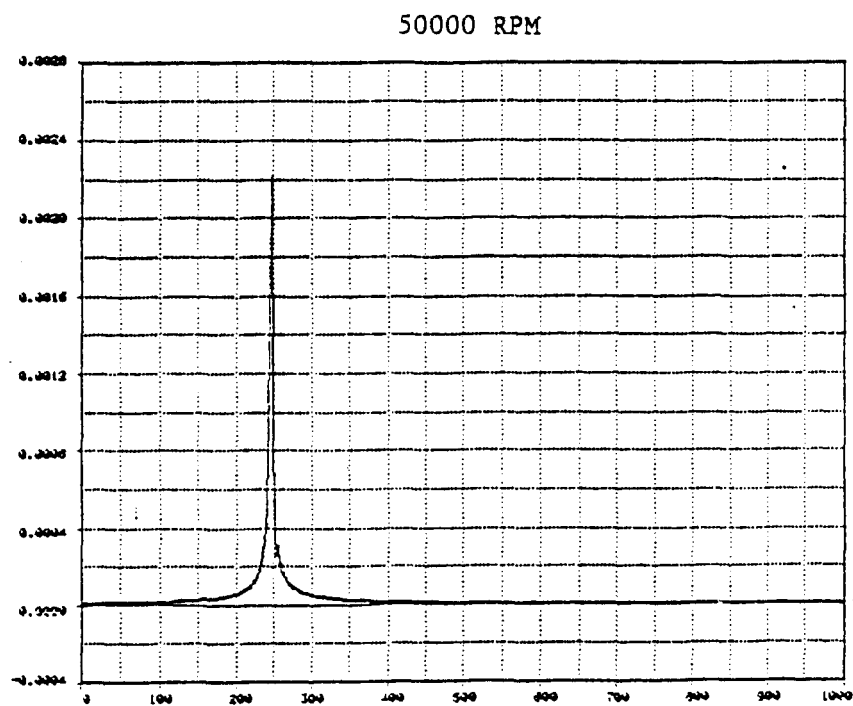
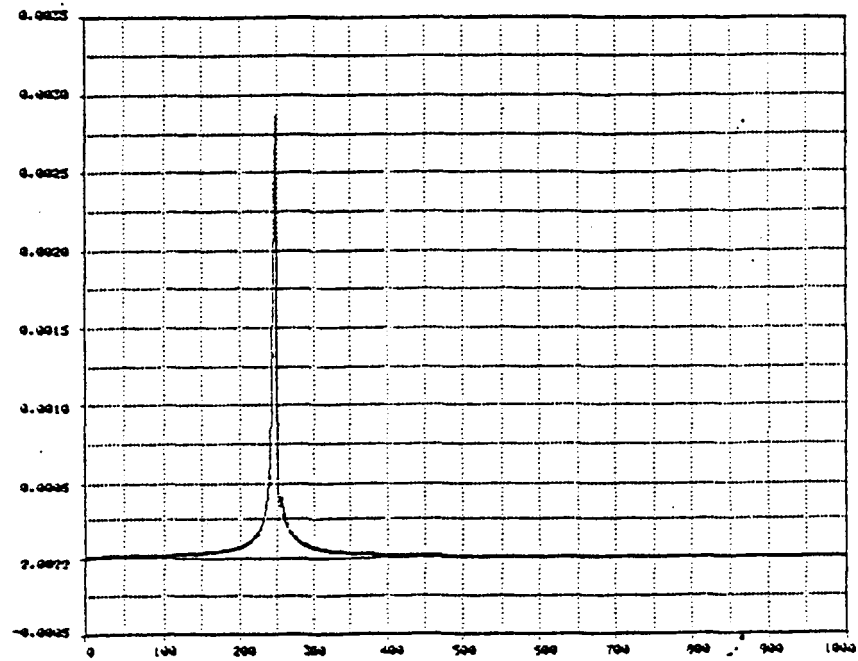


Figure 4

52000 RPM



53000 RPM

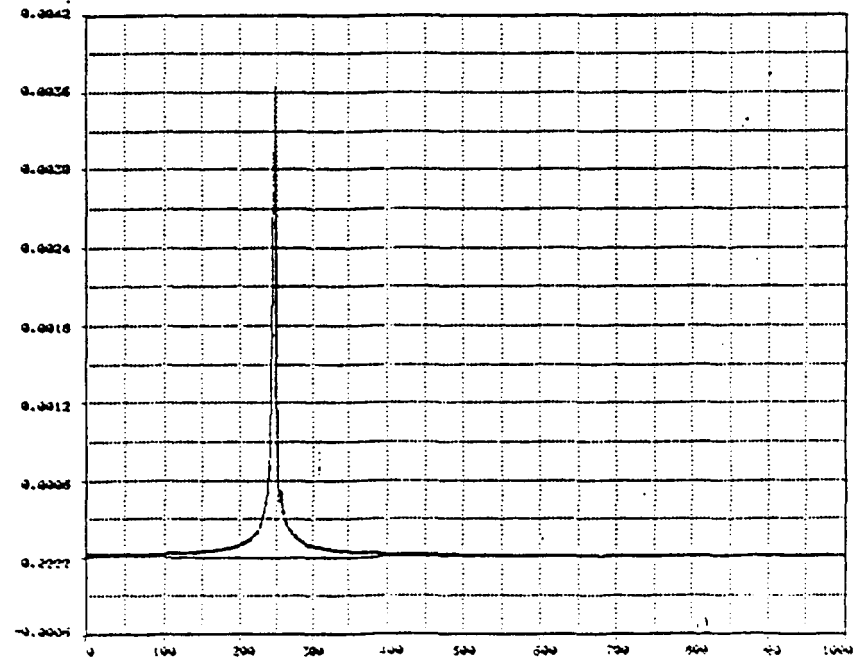


Figure 5

XXXIII-18

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

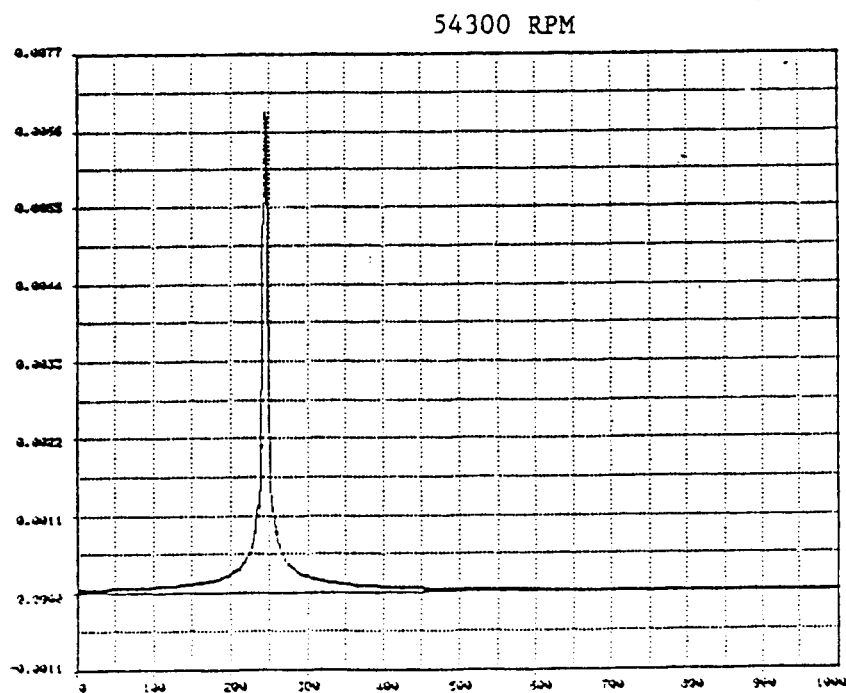
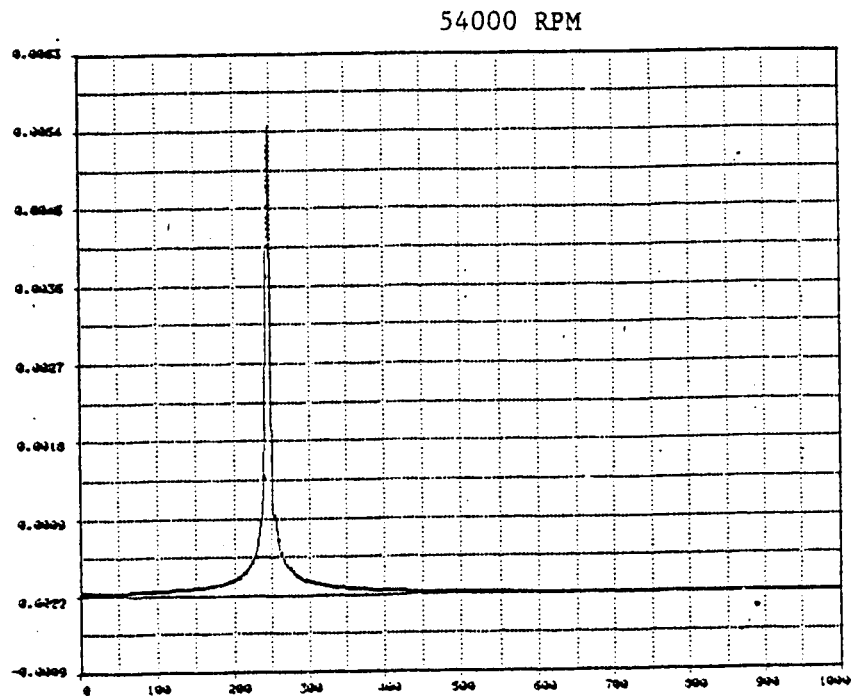
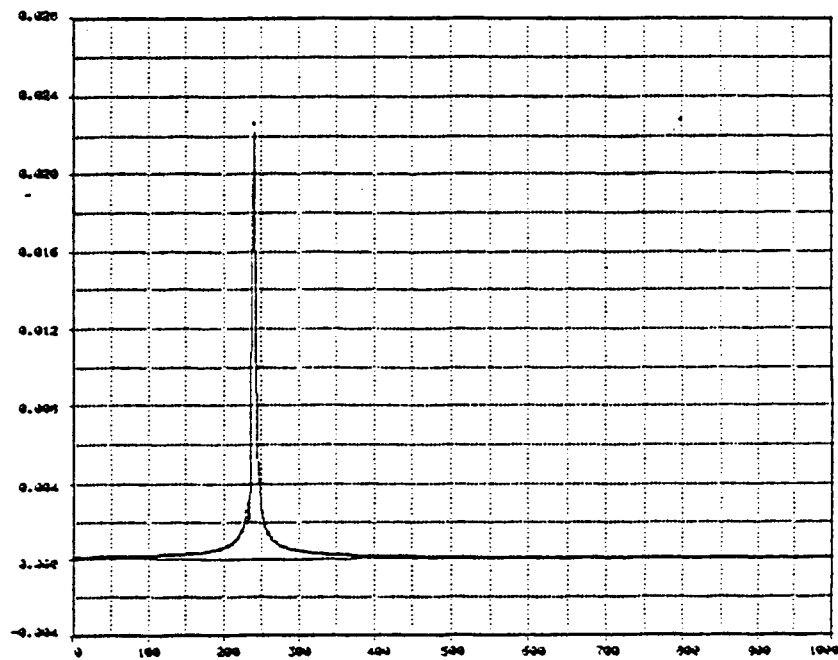


Figure 6

54600 RPM



55000 RPM

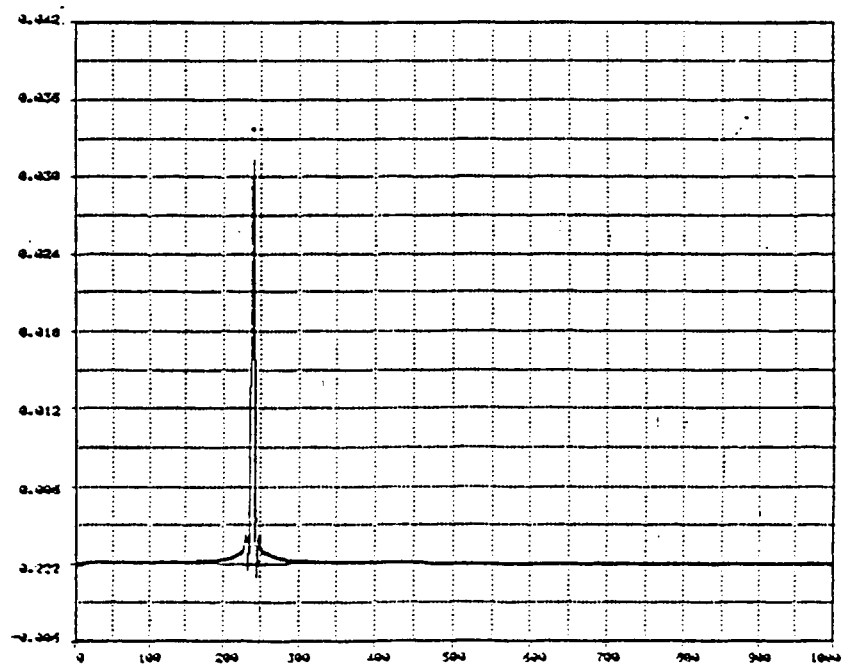
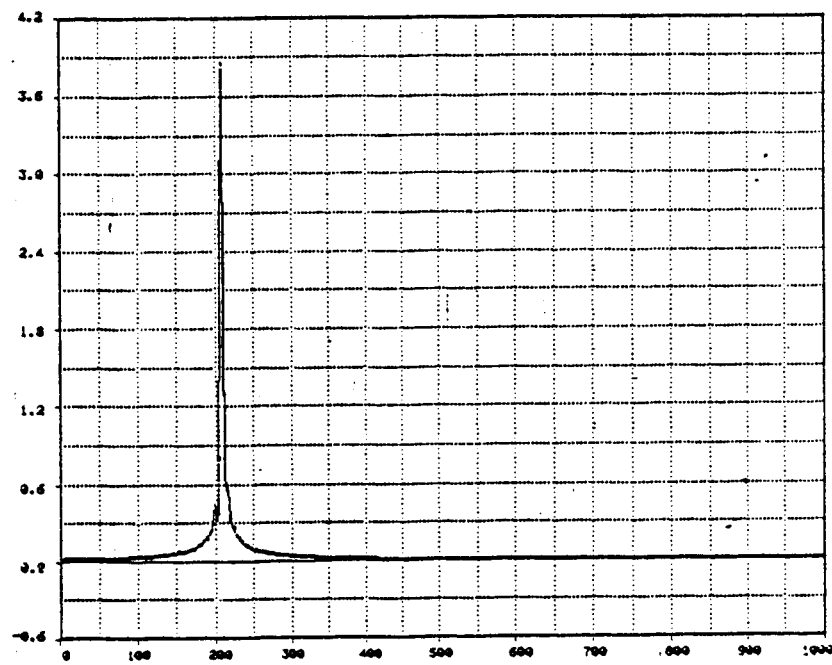


Figure 7

ORIGINAL PAGE IS
OF POOR QUALITY

56000 RPM



57000 RPM

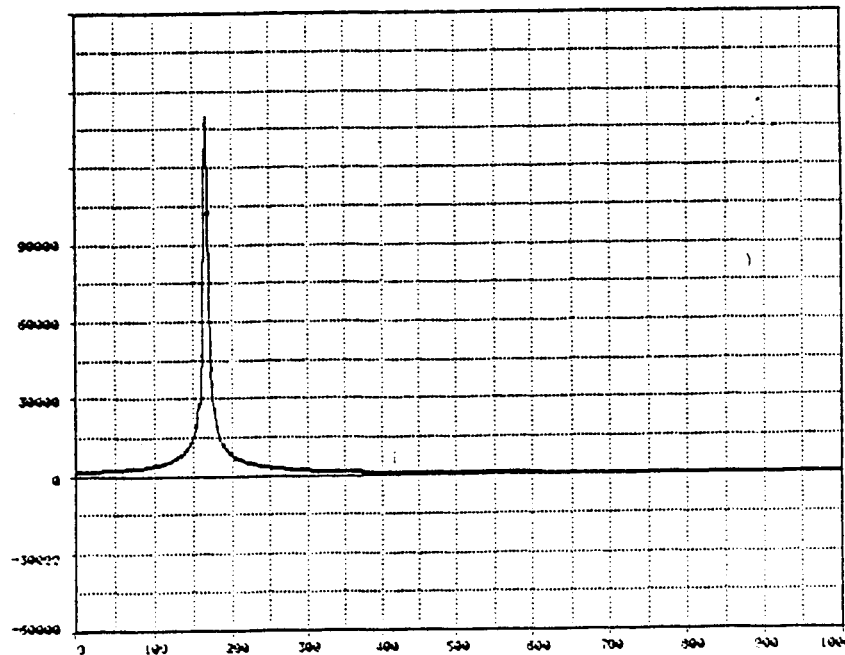


Figure 8

N89-21760

1988

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

SPACE SHUTTLE MAIN ENGINE NUMERICAL MODELING CODE
MODIFICATIONS AND ANALYSIS

Prepared by:	John P. Ziebarth
Academic Rank:	Assistant Professor
University and Department:	University of Alabama in Huntsville Computer Science Department
NASA/MSFC:	
Laboratory:	Structures and Dynamics
Division:	Aerophysics
Branch:	Computational Fluid Dynamics
MSFC Colleagues:	Luke Schutzenhofer and Paul McConnaughey
Date:	August 17, 1988
Contract Number:	NGT 01-002-099 The University of Alabama

SPACE SHUTTLE MAIN ENGINE
NUMERICAL MODELING CODE
MODIFICATIONS AND ANALYSIS

by

John P. Ziebarth
Assistant Professor of Computer Science
University of Alabama in Huntsville
Huntsville, Alabama 35899

ABSTRACT

The user of Computational Fluid Dynamics (CFD) codes must be concerned with the accuracy and efficiency of the codes if they are to be used for timely design and analysis of complicated three-dimensional fluid flow configurations. A brief discussion of how accuracy and efficiency affect the CFD solution process is given. A more detailed discussion of how efficiency can be enhanced by using a few Cray Research Inc. utilities to address vectorization is presented and these utilities are applied to a three-dimensional Navier-Stokes CFD code (INS3D).

ACKNOWLEDGMENTS

I am grateful for the opportunity that I have had to participate in the Summer Faculty Fellowship Program. The work that I completed this summer would not have been possible without the advice and assistance of Luke Schutzenhofer and Paul McConnaughey.

I also appreciate the support of Dr. Carl Davis and the University of Alabama in Huntsville for allowing me to work with this program.

I would like to thank Dr. Mike Freeman for his encouragement and guidance throughout the program. The experience at Marshall Space Flight Center has been very enjoyable, and I am indebted to the National Aeronautics and Space Administration and the American Society for Engineering Education for providing me with the chance to work at NASA/Marshall Space Flight Center in Huntsville, Alabama.

LIST OF FIGURES

Figure 1	Computer Requirements for Computational Aerodynamics	XXXIV - 6
Figure 2	Loopmark Results for Original Version of Subroutine VISRHS2	XXXIV - 11
Figure 3	Loopmark Results for Modified Version of Subroutine VISRHS2	XXXIV - 12
Figure 4	Flowtrace Summary for Original Version of INS3D	XXXIV - 13
Figure 5	Flowtrace Summary for Modified Version of INS3D	XXXIV - 14
Figure 6	Flowtrace Calling Tree for INS3D	XXXIV - 15

INTRODUCTION

Computational Fluid Dynamics (CFD) has become an extensively used tool in the design and analysis of complex three-dimensional flows including those through the Space Shuttle Main Engine (SSME). Rapid advances in CFD over the last decade have provided to the user community a large set of computer codes, each with various capabilities and constraints. These codes are adopted for use by scientists and engineers who may or may not understand fully the physics and mathematics in the code. Once adopted, these codes are adapted to solve a variety of problems, which hopefully, are similar enough to the ones for which the code was written to be applicable to the physics in the code.

In addition to this accuracy consideration, is the concept of efficiency. For full three-dimensional calculations, the amount of supercomputer time necessary to reach a solution can be quite large. To code developers this large computational time may not be of much concern; however, for code users interested in design and/or analysis, solutions must be achievable in a reasonable amount of time. Unfortunately, most current university curricula for engineers and scientists do not contain any preparation on how to efficiently write FORTRAN or use current supercomputers.

How to measure and/or evaluate accuracy and efficiency is of current interest to the CFD community in general, and to the CFD Branch at NASA/MSFC in particular. In fact, the entire process of doing a CFD calculation is somehow affected by either accuracy or efficiency. The solution process can be divided into three major components: pre-processing, processing and post-processing. Pre-processing involves primarily the geometry modeling and the grid generation. Processing involves the equation modeling of the flow physics, the implementation of a numerical method to solve the equations, the FORTRAN coding of the numerical method and the running of the computer program. The post-processing takes the massive amount of data produced by the code and transforms it into a usable form by some visualization methods.

OBJECTIVES

The objectives of this work were to:

- 1) Consider accuracy and efficiency as it relates to CFD codes, and begin to evaluate and establish guidelines and criteria for CFD code users.
- 2) Consider efficiency with respect to reducing the amount of CPU time needed to reach a solution. Do this by applying some Cray Research Inc. utilities to a three-dimensional Navier-Stokes CFD code (INS3D).

ACCURACY

Accuracy is affected at all the levels of processing. In the pre-processing phase, a precise modeling of the geometry is necessary. Regardless of how this geometry modeling is done it quite often only approximates some complicated boundaries and corners of three-dimensional configurations. The grid generation is then done on the resulting geometry model by some mathematical method. It is well documented [1, 2] that the final solution is affected by the distribution of grid points in this computational domain; however, quantitative measures of how "good" a grid is are not readily available.

In the processing phase of the solution accuracy is affected by many factors. Although the Navier-Stokes equations are generally accepted as a full description of turbulent fluid motion in a continuum, the complexity of the equations and the extremely small time and length scales of turbulent motion prohibit practical numerical computation of turbulent flows by this method. Thus, many levels of approximation are used. These include both linear and nonlinear inviscid, boundary layer, Reynolds averaged Navier-Stokes and large eddy simulation approximations. Once a set of equations is chosen to solve the flow physics, a numerical method to solve these equations must be implemented. Thus, accuracy has been affected at two levels in the processing so far. The next step is the coding (typically in FORTRAN) and the running of the code. Here accuracy is dependent on correct coding, precision of the computer system, and degree to which the solution is allowed to converge.

In the post-processing, accuracy can be affected by the visualization method used. Visualization of the flow involves taking the computed data and inputting it into a software package which can be either a commercial or locally developed product and outputting the results onto some type of graphics hardware. Thus, correctly written software is required and the resolution of the output device medium makes a difference in the accuracy of the visualization.

Currently those involved in CFD approach accuracy from different directions. Some accept the "answers" as produced by existing codes as being reasonable while others tend to be skeptical of at least some aspects of the solution process. Probably the most accepted components

are the geometry modeling and the post-processing. The grid is often thought to be acceptable if a flow solution can be arrived at by using it. Most doubt, if any exists, is usually directed at the flow solution itself. Unfortunately, questions concerning the accuracy of the flow solution do not necessarily have simple solutions. Many factors play a part in the solution; the modeling of the equations, the numerical method used to solve these equations, the convergence criteria used, boundary condition implementation, turbulence modeling, grid dependencies, correct coding, etc.

EFFICIENCY

The efficiency of a CFD solution also affects pre-processing, processing, and post-processing. This is especially true in design/analysis environments and in situations where computational resources are scarce. Pre-processing has typically been the most time consuming part of the solution process, especially when an analysis is being done on a new configuration. Current efforts in interactive geometry modeling and grid generation have helped some, but this phase is still an area of ongoing research and development. In the processing phase efficiency is tightly coupled to computational resources. An interesting projection [3] was made which indicates that to be useful in design, computational aerodynamics requires machines capable of at least one trillion floating-point operations per second (see Fig. 1). Current supercomputers do not yet meet this requirement. Post-processing has not been a particularly inefficient part of the entire solution process if all that is required is inputting the solution values to a software package and then displaying the results on a graphics workstation. However, this is a time consuming operation if real time high resolution graphical animation of the flow is desired. This can be a very computationally intensive task and can involve the transmission of extremely large quantities of data.

The concern of this current work is efficiency during the processing phase. We assume the pre-processing has been done in what follows and no post-processing is discussed. Current CFD users are constrained in achieving a solution by the following factors:

- a) The CPU time needed to reach an acceptable level of convergence may be large (greater than one hour).
- b) Supercomputing centers are often saturated, thus, the CPU time translates to a much larger wall clock time.
- c) Most or all of the CFD code is typically written by a researcher who often is not concerned with or perhaps even knowledgeable of how to write efficient FORTRAN code.

A CFD user can improve efficiency in various ways. One way is to modify the code so that the FORTRAN is more efficient. This is typically referred to as optimization and is not related to vectorization. Optimization improves efficiency even on a scalar processor. Although writing

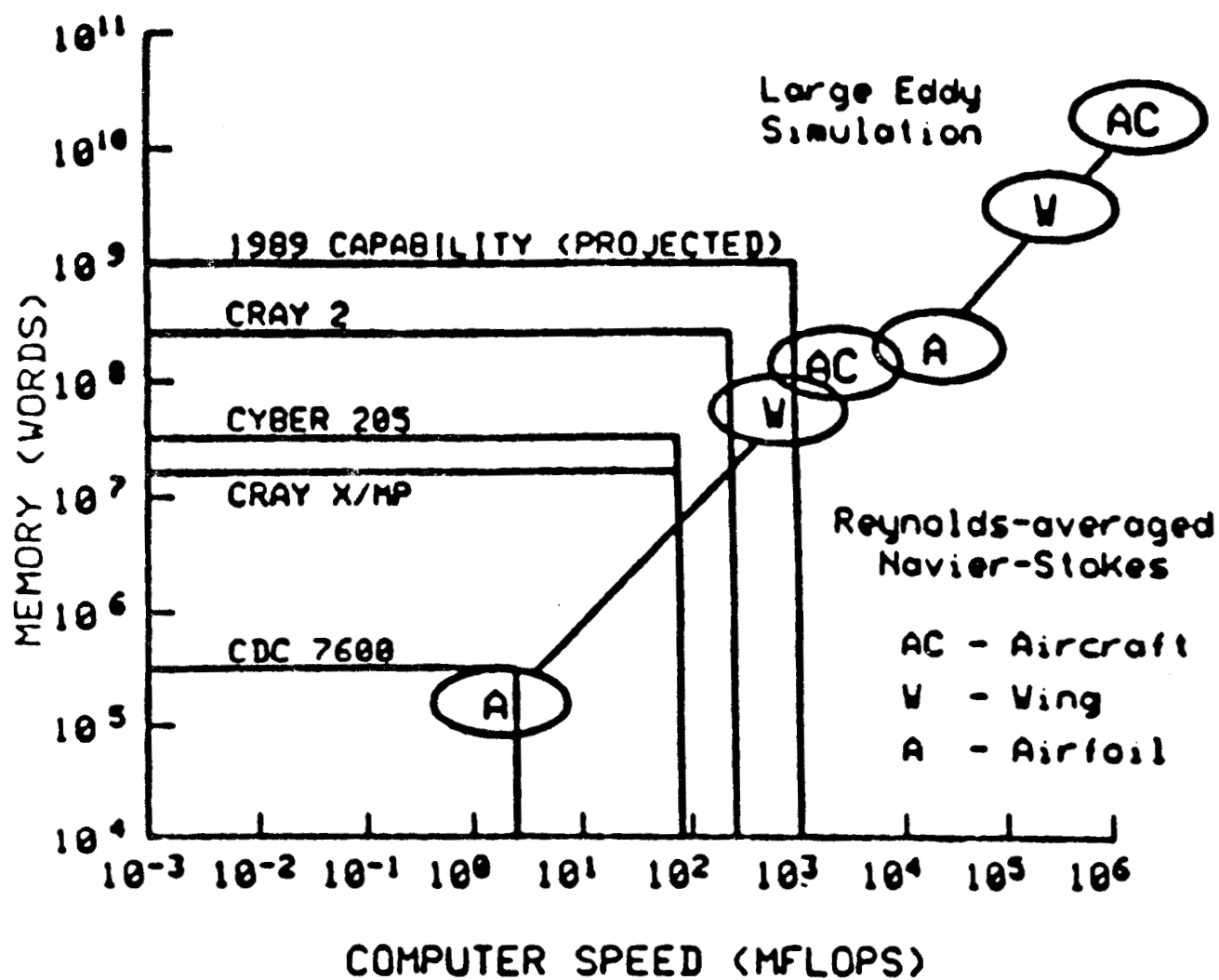


Figure 1 Computer Requirements for Computational Aerodynamics

efficient optimized code is best done during the initial code development, it is usually neglected at that time for two reasons; 1) getting a solution is of a higher priority than is writing efficient code and 2) many CFD researchers do not know how to write efficient FORTRAN code. For the user, optimizing the code can be very time consuming, requiring a major rewriting of much of the code. Methods for optimization are discussed in [4].

Vectorization is a capability of current supercomputers which yields the greatest efficiency benefit. It essentially is parallel processing since it implies that a single instruction performs many operations instead of just one; however, in the context of this report, parallel processing and vectorization will be discussed separately. More information on vectorization can be found in [4].

Fortunately for CFD users vectorization is a capability automatically implemented by the FORTRAN compilers of current supercomputers. These compilers have matured over the last few years to the point where they do a pretty good job, and they should continue to improve with time. Most CFD users currently depend on this automatic capability and actually never know which loops (FORTRAN DO loops) actually do or do not vectorize. Two capabilities of the Cray FORTRAN compiler will be discussed and their effect on a three-dimensional incompressible Navier-Stokes code will be described.

Before continuing with the vectorization discussion a few comments on parallel processing are appropriate. Parallel processing, also called multitasking (at least by Cray Research Inc.), is the capability to divide the solution into segments such that they are being done concurrently on more than one CPU. Multitasking is further divided into two parts, macrotasking and microtasking. Macrotasking refers to executing multiple segments of a program simultaneously through library calls. This requires restructuring of the code by the user and is generally not a trivial task. Microtasking refers to being able to simultaneously execute segments of a program at the DO loop level through compiler directives. This is generally easier to do than macrotasking, creates less overhead, and produces a code which from the computer system management point of view is very nice because it runs even if only one CPU is available. If and when another CPU becomes available the microtasked code has the ability to make use of the free CPU (or CPUs). Implementing microtasking is important for the CFD user although typically it is not done. This is primarily because it is a relatively new feature and little effort has been made to train CFD users to make use of it.

Fortunately for users, compilers in the future will be able to do at least some microtasking automatically. Cray should eventually release a feature called autotasking, which will automatically produce multitasked code for certain program structures.

An important point for users to be aware of is that multitasking reduces wall clock time, whereas vectorization and optimization reduce CPU clock time. The user should always reduce CPU time first by all methods available then reduce wall clock time through multitasking.

VECTORIZATION TOOLS AND INS3D

Two of the UNICOS utilities available from Cray Research Inc. for their Cray X-MP computer system are loopmark and flowtrace. These utilities are available under Cray's CFT77 compiler. A warning is in order for users of Cray computer systems. Cray offers for their X-MP systems both COS and UNICOS operating systems and various compilers (or versions of compilers). Consult the manuals for the system being used to see which utilities are available and how to implement them. Also, over the last decade compilers have continued to get "smarter", so FORTRAN code may run differently and vectorization may be applied differently now than in the past. Be aware of this when comparing results.

The loopmark and flowtrace utilities will be applied and results discussed for INS3D. The geometrical configuration considered is three-dimensional flow past a circular cylinder between two parallel plates. See [5] for a discussion of this case.

The loopmark utility can be used to determine which inner DO loops were automatically vectorized by the compiler. The listing will bracket the DO loops and will indicate which vectorized and which did not (see Figs. 2 and 3). Loopmark also furnishes a reason for the loop not vectorizing. Knowledge of what causes and what inhibits vectorization [4] is then necessary so that the code can be modified to try and implement vectorization in loops where the compiler could not. This involves a restructuring of the code. Care must be taken to be sure the code runs the same way and yields identical results before and after the restructuring. This typically means only a few changes should be implemented between check runs of the code.

Loopmark creates a listing file which may be much longer than the original code. Figures 2 and 3 are only small portions of the original and modified subroutine VISRHS2 and are only intended to show the structure of the listing. The INS3D code is approximately 5570 lines long and subroutine VISRHS2 is 650 lines long, so Figures 2 and 3 only show the format of the output. The left column in each figure shows sequential line numbers with two dots inserted for missing code. Figure 2 shows a portion of VISRHS2 where three nested DO loops begin. The DO 50 loop is a candidate for vectorization. Figure 3 shows VISRHS2 restructured so the original inner DO loop (DO 50 in Fig.

2) now is broken into four shorter DO loops (DO 50, DO 51, DO 52 and DO 53), and each of these loops does vectorize. The 50 CONTINUE in Figure 2 and the 53 CONTINUE in Figure 3 represent the same location in the code. Also included in Figures 2 and 3 is the vectorization information which is a part of the output of loopmark.

Flowtrace is a utility which will monitor calls to and from routines in the code and print various statistics about total execution time (see Figs. 4 and 5). In Figures 4 and 5 the columns from left to right are the routines called, the execution time in seconds for each, the percentage that each routine used of the total execution time, the number of times the routine was called, the average time (number of times called divided by execution time for routine) and finally the calling program unit for each routine. Flowtrace also prints out a calling tree (Fig. 6) for the program. The indentations of the routines in the tree indicate the levels of depth in the tree.

Flowtrace does add overhead to the program run but it is a very useful utility. It can also be enabled only for parts of a program thus creating less total overhead. When using flowtrace pay attention to the percent column. Man hours (or days) should probably not be spent rewriting and restructuring sections of the code where little time is spent. Also realize that as a certain section of the code is speeded up, its execution time and percentage of total time will decrease, but the percentage of another section of the code will increase since the percentages must always sum to 100.

Figure 4 indicates that subroutine VISRHS2 consumes most of the CPU time and this represents the largest percentage of total time relative to any other subroutine. By looking first at the flowtrace output and then at the compiled code with loopmark enabled (Fig. 2) it is noted that only a small portion of the DO loops in VISRHS2 vectorize automatically. By restructuring of the code all inner DO loops in VISRHS2 can be vectorized and now VISRHS2 consumes 24.3 percent (Fig. 5) of the total time rather than the 36.2 percent seen in Figure 4 and the execution time for the subroutine is reduced from 690 seconds to 389 seconds. Thus, by simply restructuring one subroutine a speedup of 1.2 times can be attained. The subroutine VISRHS2 which is now vectorized still requires the most time and thus is a good candidate for microtasking of the triply nested DO loops in it.

ORIGINAL PAGE IS
OF POOR QUALITY

```

4068      3.          C*****
4069      4.          SUBROUTINE VISRHS2
4070      5.          C*****

4107      42.          C
4108      43. S-----<      DO 100 K=KEND2,KENDM
4109      44. :   S-----<      DO 100 L=2,LM
4110      45. :   S          C

4114      49. :   S          IKL=(K-1)*KK+(L-1)*LL
4115      50. :   S          C
4116      51. :   :   S---<      DO 50 J=1,JMAX
4117      52. :   :   S          IJL=J+(L-1)*LL
4118      53. :   :   S          I=IKL+J
4119      54. :   :   S          C-----
4120      55. :   :   S          C...CALCULATE XK, YK, ZK
4121      56. :   :   S          C-----

4228      163. :   :   S          C          A(2,3,J) = VNUTJ* (XSIX**2 + XSIY**2 + XSIZ**2)
4229      164. :   :   S          A(3,3,J) = VNUTJ* (XSIX*ETAX + XSIY*ETAY + XSIZ*ETAZ)
4230      165. :   :   S          A(4,3,J) = VNUTJ* (XSIX*ZETAX + XSIY*ZETAY + XSIZ*ZETAZ)
4231      166. :   :   S--->      50 CONTINUE
4232      167. :   S          C=====

```

VECTORIZATION INFORMATION

*** *** Loop starting at line 51 was not vectorized because
a value is defined in a conditionally executed block but used in another block of the loop

Figure 2 Loopmark Results for Original Version of
Subroutine VISRHS2

```

4068      3.      C*****
4069      4.      SUBROUTINE VISRHS2
4070      5.      C*****

4111      46. S-----<      DO 100 K=KEND2,KENDM
4112      47. :   S-----<      DO 100 L=2,LM
4113      48. :   S           C

4117      52. :   S           IKL=(K-1)*KK+(L-1)*LL
4118      53. :   S           C
4119      54. :   S           IF (KPERI .EQ. 1) then
4120      55. :   :   V---<      DO 50 J=1,JMAX
4121      56. :   :   V           IJL=J+(L-1)*LL

4130      65. :   :   V--->      50 continue
4131      66. :   S           c
4132      67. :   S           else
4133      68. :   S           c
4134      69. :   :   V---<      DO 51 J=1,JMAX
4135      70. :   :   V           IJL=J+(L-1)*LL

4147      82. :   :   V           ZKnn(j) = ( Z(IRR)- 8.*(Z(IR)-Z(IP))- Z(IPP) )/12.
4148      83. :   :   V--->      51 continue
4149      84. :   S           c
4150      85. :   S           endif
4151      86. :   S           c
4152      87. :   :   V---<      DO 52 j=1,JMAX
4153      88. :   :   V           IJL=J+(L-1)*LL

4176     111. :   :   V           endif
4177     112. :   :   V--->      52 continue
4178     113. :   S           c
4179     114. :   :   V---<      do 53 j = 1,jmax
4180     115. :   :   V           IJL=J+(L-1)*LL

4218     153. :   :   V           A(4,3,J) = VNUTJ* (XSIX*ZETAX + XSIY*ZETAY + XSIZ*ZETAZ)
4219     154. :   :   V--->      53 continue
4220     155. :   S           C
4221     156. :   S           C-----

```

VECTORIZATION INFORMATION

```

*** *** Loop starting at line 55 was vectorized
*** *** Loop starting at line 69 was vectorized
*** *** Loop starting at line 87 was vectorized
*** *** Loop starting at line 114 was vectorized

```

Figure 3 Loopmark Results for Modified Version of Subroutine VISRHS2

ORIGINAL PAGE IS
OF POOR QUALITY

1 FLOWTRACE -- Alphabetized summary

0	Routine	Time executing	Called	Average T	
28	BC	2.953	(0.15%) 800	0.005	@02534712a Called by MAIN
14	COMET	102.060	(5.36%) 4453200	>	@02540487a Called by STEP
20	ETAINV	47.878	(2.50%) 19800	0.002	@02561161a Called by STEP
9	FLUXVE	46.220	(2.43%) 2226800	>	@02541718a Called by RHS
3	GRID	0.016	(0.00%) 1	0.016	@02541775a Called by INITIA
5	IC	0.008	(0.00%) 1	0.008	@02542222a Called by INITIA
2	INITIA	>	(0.00%) 1	>	@02542313a Called by MAIN
4	JACOB	0.257	(0.01%) 1	0.257	@02542452a Called by INITIA
1	MAIN	0.025	(0.00%) 1	0.025	@02534546a Called by
8	METRIC	111.373	(5.84%) 4453200	>	Called by RHS STEP
	@02543141a				2226800 2226800
28	OUTPUT	0.032	(0.00%) 12	0.003	@02545674a Called by MAIN
7	RHS	129.922	(6.82%) 800	0.217	@02552200a Called by STEP
12	SMOOTH	60.761	(3.19%) 800	0.101	@02552772a Called by RHS
6	STEP	213.562	(11.21%) 800	0.356	@02553733a Called by MAIN
19	TK	122.022	(6.40%) 2226800	>	@02554778a Called by STEP
15	TKINV	159.074	(8.35%) 2226800	>	@02555201a Called by STEP
18	TRI	14.113	(0.74%) 72000	>	Called by XIINV ZETAINV
	@02555457a				32400 39600
17	TRI2	10.001	(0.52%) 36000	>	Called by XIINV ZETAINV
	@02555571a				16200 19800
22	TRIP	14.930	(0.78%) 39600	>	@02555720a Called by ETAINV
21	TRIP2	10.346	(0.54%) 19800	>	@02556150a Called by ETAINV
27	VISCT	0.007	(0.00%) 800	>	@02556434a Called by MAIN
10	VISRHS	86.818	(4.56%) 2226800	>	@02556442a Called by RHS
11	VISRHS2	690.153	(36.21%) 800	1.150	@02556621a Called by RHS
16	XIINV	41.489	(2.18%) 16200	0.003	@02560350a Called by STEP
23	ZETAINV	41.989	(2.20%) 19800	0.002	@02562424a Called by STEP
*	* * * TOTAL	1905.804	18039617 Total calls		

Figure 4 Flowtrace Summary for Original Version of INS3D

```

1 FLOWTRACE -- Alphabetized summary
O Routine Time executing Called Average T
28 BC 2.950 ( 0.18%) 800 0.005 @02542883a Called by MAIN
14 COMET 101.112 ( 8.32%) 4453200 > @02548440a Called by STEP
20 ETAINV 47.241 ( 2.95%) 19800 0.002 @02570143a Called by STEP
9 FLUXVE 46.727 ( 2.92%) 2226800 > @02547887a Called by RHS
3 GRID 0.015 ( 0.00%) 1 0.015 @02547746a Called by INITIA
5 IC 0.005 ( 0.00%) 1 0.005 @02550173a Called by INITIA
2 INITIA > ( 0.00%) 1 > @02550284a Called by MAIN
4 JACOB 0.257 ( 0.02%) 1 0.257 @02550423a Called by INITIA
1 MAIN 0.025 ( 0.00%) 1 0.025 @02542517a Called by
8 METRIC 111.038 ( 8.94%) 4453200 > Called by RHS STEP
@02551112a 2226800 2226800
28 OUTPUT 0.033 ( 0.00%) 12 0.003 @02553845a Called by MAIN
7 RHS 129.301 ( 8.09%) 800 0.216 @02580151a Called by STEP
12 SMOOTH 80.290 ( 3.77%) 800 0.100 @02580743a Called by RHS
6 STEP 212.457 ( 13.29%) 800 0.354 @02581704a Called by MAIN
19 TK 121.543 ( 7.80%) 2226800 > @02582747a Called by STEP
15 TKINV 158.434 ( 9.91%) 2226800 > @02583152a Called by STEP
18 TRI 14.025 ( 0.88%) 72000 > Called by XIINV ZETAINV
@02583430a 32400 39600
17 TRI2 9.905 ( 0.62%) 36000 > Called by XIINV ZETAINV
@02583542a 16200 19800
22 TRIP 14.784 ( 0.92%) 39600 > @02583871a Called by ETAINV
21 TRIP2 10.240 ( 0.64%) 19800 > @02584121a Called by ETAINV
27 VISCT 0.007 ( 0.00%) 800 > @02584405a Called by MAIN
10 VISRHS 88.592 ( 5.42%) 2226800 > @02584413a Called by RHS
11 VISRHS2 388.858 ( 24.32%) 800 0.648 @02584572a Called by RHS
16 XIINV 41.285 ( 2.58%) 16200 0.003 @02587332a Called by STEP
23 ZETAINV 41.782 ( 2.61%) 19800 0.002 @02571408a Called by STEP
* * * TOTAL 1598.885 18039817 Total calls

```

Figure 5 Flowtrace Summary for Modified Version of INS3D

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

1 FLOWTRACE -- Calling tree

1	MAIN	02542517a
2	INITIA	02550264a
3	GRID	02547746a
4	JACOB	02550423a
5	IC	02550173a
6	STEP	02561704a
7	RHS	02560151a
8	METRIC	02551112a
9	FLUXVE	02547667a
10	VISRHS	02564413a
11	VISRHS2	02564572a
12	SMOOTH	02560743a
13	METRIC	02551112a
14	COMET	02546440a
15	TKINV	02563152a
16	XIINV	02567332a
17	TRI2	02563542a
18	TRI	02563430a
19	TK	02562747a
20	ETAINV	02570143a
21	TRIP2	02564121a
22	TRIP	02563671a
23	ZETAINV	02571406a
24	TRI2	02563542a
25	TRI	02563430a
26	BC	02542663a
27	VISCT	02564405a
28	OUTPUT	02553645a
STOP		in MAIN

Figure 6 Flowtrace Calling Tree for INS3D

CONCLUSIONS AND RECOMMENDATIONS

During the past decade CFD codes have been accepted with justifiable skepticism by users interested in design and analysis. These users are not interested in getting just a "number" out of these codes, they are interested in getting a believable "number" in an affordable amount of time. Today, with many codes promising good results, users must have criteria they can rely on to evaluate codes. They also must be able to efficiently apply these codes to various problems on often scarce or expensive computing resources.

CFD users need to be able to evaluate codes in a reasonable amount of time and with a high degree of assurance that the chosen codes are efficient and yield accurate results. Evaluation criteria should be established for this purpose. Along with establishing criteria to evaluate codes is the need to educate code users of the utilities and methods to develop efficient and accurate codes.

The application of the loopmark and flowtrace utilities to a CFD code show that efficiency can be enhanced by using basic tools available to but not necessarily known by users. Along with the development of evaluation criteria should be the education of users on how to develop efficient well-written FORTRAN code.

Besides loopmark and flowtrace are other utilities which can enhance performance of codes. Also, multitasking (parallel processing) needs to be used by CFD code users to make the most efficient use of supercomputer resources. Work is under way to apply both microtasking and macrotasking to INS3D; hopefully this work will be of interest to the user community.

REFERENCES

1. Thompson, J.F., Warsi, Z.U.A., and Mastin, C.W. Numerical Grid Generation: Foundations and Applications, North Holland, 1985.
2. Wang, T.S., Soni, B.K., "Goodness-of-Grid: Quantitative Measures", AIAA/ASME/SIAM/APS First National Fluid Dynamics Congress, July 25-28, 1988, Cincinnati, Ohio.
3. Peterson, V.L., "The Impact of Supercomputers on the Aerospace Sciences, "AIAA Twenty-fourth Aerospace Sciences Meeting, Reno, Nevada, January, 1986.
4. Ziebarth, J.P., "Introduction to Supercomputing", course notes, University of Alabama in Huntsville, 1988.
5. Rogers, S.E., Kwak, D., and Chang, J.L.C., "INS3D - An Incompressible Navier-Stokes Code in Generalized Three-Dimensional Coordinates", NASA Technical Memorandum 100012, November, 1987.